Quantum Spectrum Testing

Ryan O'Donnell* John Wright*

January 22, 2015

Abstract

In this work, we study the problem of testing properties of the spectrum of a mixed quantum state. Here one is given n copies of a mixed state $\rho \in \mathbb{C}^{d \times d}$ and the goal is to distinguish (with high probability) whether ρ's spectrum satisfies some property $\mathcal{P}$ or whether it is at least ϵ-far in ℓ_1-distance from satisfying $\mathcal{P}$. This problem was promoted in the survey of Montanaro and de Wolf [MdW13] under the name of testing unitarily invariant properties of mixed states. It is the natural quantum analogue of the classical problem of testing symmetric properties of probability distributions.

Unlike property testing probability distributions—where one generally hopes for algorithms with sample complexity that is sublinear in the domain size—here the hope is for algorithms with *subquadratic* copy complexity in the dimension d. This is because the (frequently rediscovered) "empirical Young diagram (EYD) algorithm" [ARS88, KW01, HM02, CM06] can estimate the spectrum of any mixed state up to ϵ-accuracy using only $\widetilde{O}(d^2/\epsilon^2)$ copies. In this work, we show that given a mixed state $\rho \in \mathbb{C}^{d \times d}$:

- $\Theta(d/\epsilon^2)$ copies are necessary and sufficient to test whether ρ is the maximally mixed state, i.e., has spectrum $(\frac{1}{d}, \dots, \frac{1}{d})$. This can be viewed as the quantum analogue of Paninski [Pan08]'s sharp bounds for classical uniformity-testing.
- $\Theta(r^2/\epsilon)$ copies are necessary and sufficient to test with one-sided error whether ρ has rank r, i.e., has at most r nonzero eigenvalues. For two-sided error, a lower bound of $\Omega(r/\epsilon)$ copies holds.
- $\widetilde{\Theta}(r^2)$ copies are necessary and sufficient to distinguish whether ρ is maximally mixed on an r-dimensional or an $(r+1)$-dimensional subspace. More generally, for r vs. $r+\Delta$ (with $1 \leq \Delta \leq r$), $\widetilde{\Theta}(r^2/\Delta)$ copies are necessary and sufficient.
- The EYD algorithm requires $\Omega(d^2/\epsilon^2)$ copies to estimate the spectrum of ρ up to ϵ-accuracy, nearly matching the known upper bound. In addition, we simplify part of the proof of the $\widetilde{O}(d^2/\epsilon^2)$ upper bound.

Our techniques involve the asymptotic representation theory of the symmetric group; in particular Kerov's algebra of polynomial functions on Young diagrams.

*Department of Computer Science, Carnegie Mellon University. Some of this work performed while the first-named author was at the Boğaziçi University Computer Engineering Department, supported by Marie Curie International Incoming Fellowship project number 626373. Supported also by NSF grants CCF-0747250 and CCF-1116594. The second-named author is also supported by a Simons Fellowship in Theoretical Computer Science and completed some of this work while visiting Columbia University. {odonnell,jswright}@cs.cmu.edu

Contents

1 Introduction

A common scenario in quantum mechanics involves an experimental apparatus which outputs a particle whose state is a random variable. For example, in a version of the the famous Stern–Gerlach experiment by Phipps and Taylor [PT27], the experimental apparatus produced a hydrogen atom whose electron was either in state $\left|+\frac{1}{2}\right\rangle$ or $\left|-\frac{1}{2}\right\rangle$, each with probability $\frac{1}{2}$. More generally, one can describe the output of such an apparatus as falling in an orthonormal set of states $|\Psi_1\rangle, \ldots, |\Psi_d\rangle \in \mathbb{C}^d$, distributed according to a probability distribution $\mathcal{D} = (p_1, \ldots, p_d)$. Such an object is called a *mixed state* and is often conveniently represented using the *density matrix* $\rho = \sum p_i \cdot |\Psi_i\rangle\langle\Psi_i|$. The numbers $p_1, \ldots, p_d$ are called the *spectrum* of ρ.

Given such an apparatus, a fundamental task—known as *quantum state tomography*—is to produce an estimate $\widetilde{\rho} \in \mathbb{C}^{d \times d}$ which well-approximates ρ according to some distance measure (typically, the trace distance). To do this, one repeatedly runs the apparatus to produce many (say, n) independent copies of ρ and then one processes some measurement of $\rho^{\otimes n}$ to produce an estimate $\widetilde{\rho}$. It is known [FGLE12, Footnote 2] that $O(d^4 \log(d)/\epsilon^2)$ copies of ρ are sufficient to output an estimate which is ϵ-close to ρ in the trace distance. Unfortunately, the quartic dependence on d can be prohibitively large, even for quite reasonable values of d; further exacerbating this is the fact that many quantum systems are formed as the tensor product of many smaller subsystems, in which case d is exponential in the number of subsystems.

One potential way around this problem is to note that if our actual goal in producing $\widetilde{\rho}$ is to determine whether ρ satisfies some property (e.g., is maximally mixed, has low rank, etc.), then our estimate $\widetilde{\rho}$ may be giving us far more information than we need. Thus, we can possibly test whether ρ has the property in question using a much smaller number of copies. This is the motivation behind the model of *property testing of mixed states*, as promoted in the recent survey of Montanaro and de Wolf [MdW13]. Formally, we have following definition:

Definition 1.1. A property of mixed states $\mathcal{P}$ is testable with $f(d, \epsilon)$ copies if for every $d \geq 2, \epsilon > 0$ there is an algorithm $\mathcal{T}$ which, when given $f(d, \epsilon)$ copies of a mixed state $\rho \in \mathbb{C}^{d \times d}$, behaves as follows:

- If ρ satisfies $\mathcal{P}$, then $\mathbf{Pr}[\mathcal{T} \text{ accepts}] \geq 2/3$. ("Completeness")
- If ρ is ϵ-far in trace distance from all ρ' satisfying $\mathcal{P}$, then $\mathbf{Pr}[\mathcal{T} \text{ rejects}] \geq 2/3$. ("Soundness")

The choice of probability $2/3$ here is essentially arbitrary, and it can be amplified to $1 - \delta$ at the expense of increasing the number of copies by a factor of $O(\log(1/\delta))$.

As mixed states are the quantum analogue of probability distributions, this model can be seen as the quantum analogue of the model of testing properties of probability distributions. We note that the problem of testing properties of mixed states has also appeared in the area of quantum algorithms. For example, the work of [CHW07] considers Graph Isomorphism algorithms which output a mixed state ρ satisfying a certain property if and only if the input graphs are isomorphic.

In this work, we focus on the problem of testing so-called *unitarily invariant* properties. These are properties $\mathcal{P}$ for which ρ satisfies $\mathcal{P}$ if and only if $U\rho U^\dagger$ satisfies $\mathcal{P}$ for every unitary matrix U. It is easy to see that whether a mixed state ρ has such a property depends only on ρ's spectrum (hence the name *quantum spectrum testing*). Many natural properties of mixed states are unitarily invariant, such as being the maximally mixed state, having low rank, or having low von Neumann entropy. (An example of a natural property which is *not* unitarily invariant is the property of being equal to a fixed mixed state σ, so long as σ is not the maximally mixed state.) Though it is not immediately apparent from the definitions (we will show this in Section 2.2), the model of testing

properties of mixed states from Definition 1.1 is equivalent to the following definition in the case that the property in question is unitarily invariant.

Definition 1.2. A property of spectra $\mathcal{P}$ is testable with $f(d, \epsilon)$ copies if for every $d \geq 2, \epsilon > 0$ there is an algorithm $\mathcal{T}$ which, when given $f(d, \epsilon)$ copies of a mixed state $\rho \in \mathbb{C}^{d \times d}$ with spectrum $\eta = (\eta_1, \ldots, \eta_d)$, behaves as follows:

- If η satisfies $\mathcal{P}$, then $\mathbf{Pr}[\mathcal{T} \text{ accepts}] \geq 2/3$.
- If η is ϵ-far in total variation distance from every η' satisfying $\mathcal{P}$, then $\mathbf{Pr}[\mathcal{T} \text{ rejects}] \geq 2/3$.

The main gain in using Definition 1.2 over Definition 1.1 is that we only have to reason about a total variation distance involving η rather than a trace distance involving ρ, which is in general a more complicated distance measure. We note that the spectrum of a matrix is more properly thought of as an unordered multiset of eigenvalues rather than an ordered tuple, and therefore any property of spectra $\mathcal{P}$ by necessity depends only on the multiset of values $\{\eta_1, \ldots, \eta_d\}$ and not on their ordering. Hence, quantum spectrum testing corresponds in the classical world to the model of testing *symmetric* properties of probability distributions. As we will soon see, Definition 1.2 allows us to show a formal correspondence between these two models.

1.1 Classical property testing of probability distributions

The topic of property testing was introduced by Rubinfeld and Sudan in [RS92, RS96] in the context of testing algebraic properties of polynomials over finite fields. Since then, it has found applications in a wide variety of areas, including testing properties of graphs and of Boolean functions. Over the past fifteen years, an extremely successful branch of property testing, first explicitly defined in [BFR+00, BFR+13], has focused on testing properties of discrete probability distributions. In the model of testing properties of probability distributions, there is an unknown distribution $\mathcal{D}$ on the set $\{1, \ldots, d\}$, and the tester may draw a *random word* of length n from $\mathcal{D}^{\otimes n}$; i.e., obtain a sequence of n i.i.d. samples from $\mathcal{D}$. Its goal is to decide whether $\mathcal{D}$ has some property $\mathcal{P}$ or is ϵ-far from $\mathcal{P}$ in total variation distance, while minimizing n.

It is well known [DL01, pages 10 and 31] (cf. [Dia14, Slide 6]) that after taking $n = \Theta(d/\epsilon^2)$ samples from $\mathcal{D}$, the empirical distribution is ϵ-close to $\mathcal{D}$ with high probability. As a result, any property of probability distributions is testable with a linear (in d) number of samples; thus research in this area is directed at finding algorithms of *sublinear* sample complexity for various properties. That such algorithms could exist is suggested by the following Birthday Paradox-based fact:

Fact 1.3. *$\Theta(\sqrt{r})$ samples are necessary and sufficient to distinguish between the cases when the distribution is uniform on either r or $2r$ values. (The bound also holds for r vs. r' when $r' > 2r$.)*

Setting $r = \frac{d}{2}$, we see that this fact gives a sublinear algorithm for distinguishing between the uniform distribution and a distribution that is uniform on exactly half of the elements of $\{1, \ldots, d\}$. This fact is also important as it immediately gives a lower bound of $\Omega(\sqrt{d})$ for testing a variety of natural problems, those for which Fact 1.3 appears as a special case.

Perhaps the most basic property of probability distributions one can test for is the property of being equal to the uniform distribution, Unif_d. A $\Omega(\sqrt{d})$ lower bound follows directly from Fact 1.3. On the other hand, a $O(\sqrt{d}/\epsilon^4)$ upper bound was shown in the early work of [BFR+00, BFR+13] using techniques of [GR11]. The correct sample complexity was finally pinned down by Paninski in [Pan08], who showed matching upper and lower bounds:

Theorem 1.4 ([Pan08])**.** *$\Theta(\sqrt{d}/\epsilon^2)$ samples are necessary and sufficient to test whether $\mathcal{D}$ is the uniform distribution Unif_d.*

This result was recently extended [VV14] to an $O(\sqrt{d}/\epsilon^2)$ upper bound for testing equality to *any* fixed distribution, improving on the previously known [BFF+01] upper bound of $\widetilde{O}(\sqrt{d}/\epsilon^4)$. More precisely, [VV14] upper-bounds the sample complexity of testing equality to a fixed distribution $\mathcal{D}$ by $O(f(\mathcal{D})/\epsilon^2)$, where $f(\mathcal{D})$ is a certain norm which is maximized when $\mathcal{D}$ is the uniform distribution. Thus the uniform distribution is the hardest fixed distribution to test equality to.

The property of being the uniform distribution falls within the class of *symmetric* properties of probability distributions. These are the properties $\mathcal{P}$ for which $\mathcal{D} = (p_1, \ldots, p_d) \in \mathcal{P}$ if and only if $(p_{\pi(1)}, \ldots, p_{\pi(d)}) \in \mathcal{P}$ for every permutation π. Other interesting symmetric properties include having small entropy or small support size. Testing for small support size does not appear to have been precisely addressed in the literature; however the following is easy to derive from known results (in particular, the lower bound follows from the work of [VV11a]):

Theorem 1.5. *To test (with ϵ a constant) whether a probability distribution has support size r, $O(r)$ samples are sufficient and $\Omega(r/\log(r))$ samples are necessary.*

Let us now relate this section back to the main topic of this paper. As we saw earlier, the spectrum of a mixed state can be thought of as a probability distribution on the numbers $\{1, \ldots, d\}$ (indexing the associated eigenvectors); thus any property of mixed state spectra is simply a symmetric property of probability distributions. This correspondence allows us to directly compare the difficulty of testing properties of mixed state spectra and of probability distributions. In fact, the quantum case is always at least as difficult as the classical case; the reason is that the classical problem is equivalent to the quantum problem under the promise that the n "samples" provided are known orthogonal pure states, $|1\rangle, \ldots, |d\rangle$. Alternatively, in Sections 2.3.2 and Section 2.5 we will observe the following purely classical characterization of quantum spectrum testing:

Fact 1.6. *Let $\mathcal{P}$ be a symmetric property of probability distributions on $\{1, \ldots, d\}$. Testing whether the spectrum of a d-dimensional quantum mixed state satisfies $\mathcal{P}$ is equivalent to the following classical testing problem: Test whether a probability distribution $\mathcal{D}$ satisfies $\mathcal{P}$ when one is not allowed to see the whole random word $\boldsymbol{w} \sim \mathcal{D}^{\otimes n}$, but only the following d statistics: the length of the longest k-increasing subsequence of $\boldsymbol{w}$, for each $1 \leq k \leq d$. Here a* k-increasing subsequence *means a disjoint union of k weakly increasing subsequences.*

In light of the above remarks we record the following fact:

Fact 1.7. *Let $\mathcal{P}$ be a symmetric property of probability distributions which requires $f(d, \epsilon)$ samples to test classically. Then testing whether a mixed state's spectrum satisfies $\mathcal{P}$ also requires at least $f(d, \epsilon)$ copies of the mixed state.*

Although quantum spectrum testing is at least as hard as testing symmetric properties of probability distributions, there are some interesting nontrivial properties which have the same complexity in both models (up to constant factors). For example, if $\mathcal{P}$ is the property of having support size 1, then $\Theta(1/\epsilon)$ samples/copies are necessary and sufficient to test $\mathcal{P}$ in both models (see [MdW13] for the $O(1/\epsilon)$ quantum spectrum testing upper bound). In general, however, it is known that spectrum testing can require an asymptotically higher complexity (at least in terms of the parameter d).

We end this section by pointing out that a large portion of the property testing literature concerning entropy and support size actually considers the problems of either computing these values [Pan04, BDKR05, VV11a, VV11b] (within some tolerance) or distinguishing between the cases when these values are either large or small [Val08] (often these problems have some added guarantee on the probability distribution, such as all of its nonzero probabilities being sufficiently

large). These problems, strictly speaking, do not fit within the above property testing framework. In this work, when we consider the problem of testing a mixed state's rank (the quantum analogue of support size) we will be doing so explicitly within the property testing framework.

1.2 Related work

Returning to quantum spectrum testing, we would like to mention two prior lines of research that are directly relevant. The first is an algorithm—which we call the *empirical Young diagram* (EYD) algorithm—for learning the spectrum of an unknown mixed state. This algorithm is naturally suggested by the early work of Alicki, Rudnicki, and Sadowski [ARS88] and was explicitly proposed by Keyl and Werner [KW01]. Regarding its performance guarantee, Hayashi and Matsumoto [HM02] gave explicit error bounds and a short proof, but their work contained some small calculational errors, subsequently corrected by Christandl and Mitchison [CM06]. From the last of these it is easy to deduce the following:

Theorem 1.8. *The empirical Young diagram algorithm, when given $O(d^2/\epsilon^2 \cdot \ln(d/\epsilon))$ copies of a mixed state ρ with spectrum η, outputs with high probability an estimate of η that is ϵ-close in total variation distance.*

We will give a description of this algorithm later in the paper; for now, suffice it to say that it can be viewed as the quantum version of the natural classical algorithm for learning an unknown distribution, viz., outputting the empirical distribution. The EYD algorithm gives a near-quadratic improvement over known quantum state tomography algorithms for the problem of estimating a mixed state's spectrum.[1] As a result, testing properties of quantum spectra is easy with a quadratic number of copies, and so we hope for *subquadratic* algorithms.

The second result comes from the work of Childs et al. [CHW07]. It can be thought of as a quantum analogue of Fact 1.3:

Theorem 1.9. *$\Theta(r)$ copies of a state ρ are necessary and sufficient to distinguish between the cases when ρ's spectrum is uniform on either r or $2r$ values. (The bound also holds for r vs. cr when $c > 2$ is an integer.)*

Setting $r = \frac{d}{2}$, Theorem 1.9 gives a linear lower bound of $\Omega(d)$ for various properties of spectra. This is in contrast with property testing of probability distributions, in which sublinear algorithms are the main goal, with the Birthday Paradox typically precluding sub-$O(\sqrt{d})$-sample algorithms.

Finally, we mention that we may also obtain relevant results by applying Fact 1.7 to known lower bounds for classical property testing of probability distributions. Though in general these lower bounds are not tight, prior to our work this was (to our knowledge) the only way to produce lower bounds for testing spectra with a dependence on ϵ.

1.3 Our results

We have four main results. The first concerns the property that Montanaro and de Wolf refer to as **Mixedness**:

Theorem 1.10. *$\Theta(d/\epsilon^2)$ copies are necessary and sufficient to test whether $\rho \in \mathbb{C}^{d \times d}$ is the maximally mixed state; i.e., whether its spectrum is $\eta = (1/d, \ldots, 1/d)$.*

[1] One may note that the dependence on ϵ in Theorem 1.8 is slightly *worse* than that for full tomography; however, we speculate that this is an artifact of the analysis and that $O(d^2/\epsilon^2)$ copies suffice for the EYD algorithm.

This is the quantum analogue of Paninski's Theorem 1.4. We also remark that given the way we prove Theorem 1.10, Childs et al.'s Theorem 1.9 can be obtained as a very special case.

Our second result gives new bounds for testing whether a state has low rank.

Theorem 1.11. *$\Theta(r^2/\epsilon)$ copies are necessary and sufficient to test whether $\rho \in \mathbb{C}^{d\times d}$ has rank r with one-sided error. With two-sided error, a lower bound of $\Omega(r/\epsilon)$ holds.*

We note that the copy complexity is independent of the ambient dimension d. Knowing that a state is low rank can often make solving a given problem much simpler. For example, quantum state tomography can be made more efficient when the state is known to be low-rank [FGLE12]. Compare this to Theorem 1.5.

Next, we extend Childs et al.'s Theorem 1.9 to r vs. r' for *any* $r+1 \leq r' \leq 2r$. A qualitative difference is seen when $r' = r+1$; namely, nearly quadratically many copies are necessary.

Theorem 1.12. *Let $1 \leq \Delta \leq r$. Then $O(r^2/\Delta)$ copies are sufficient to distinguish between the cases when ρ's spectrum is uniform on either r or $r+\Delta$ eigenvalues; further, a nearly matching lower bound of $\widetilde{\Omega}(r^2/\Delta)$ copies holds.*

As above, we note that these bounds are independent of the ambient dimension d.

Our final results concern the EYD algorithm from Theorem 1.8. First, we give an arguably simpler proof of Theorem 1.8. Next, we complement this with a lower bound showing that the analysis of the EYD algorithm from Theorem 1.8 is tight up to logarithmic factors.

Theorem 1.13. *If $\rho \in \mathbb{C}^{d\times d}$ is the maximally mixed state, the algorithm from Theorem 1.8 fails to give an ϵ-accurate estimate (with high probability) unless $\Omega(d^2/\epsilon^2)$ copies are used.*

To our knowledge, no such lower bound was known previously. We remark that it is an interesting open question whether some *other* algorithm can estimate an unknown state's spectrum from a subquadratic number of copies.

1.4 Overview of our techniques

Following [ARS88, Har05, CM06, CHW07], we use techniques from representation theory of the symmetric group $\mathfrak{S}_n$. A basic tool is *Schur–Weyl duality*, which decomposes the space $(\mathbb{C}^d)^{\otimes n}$ as

$$(\mathbb{C}^d)^{\otimes n} \overset{\mathfrak{S}_n \times U_d}{\cong} \bigoplus_{\lambda \vdash n} \mathtt{P}_\lambda \otimes \mathtt{Q}_\lambda^d, \tag{1}$$

where the subspace $\mathtt{P}_\lambda$ corresponds to the symmetric group, the subspace $\mathtt{Q}_\lambda^d$ corresponds to the unitary group, and λ is a partition of n, thought of as a Young diagram. (Recall that a partition of n is a tuple $\lambda = (\lambda_1, \ldots, \lambda_\ell)$ satisfying $\lambda_1 \geq \ldots \geq \lambda_\ell \geq 0$ and $\lambda_1 + \ldots + \lambda_\ell = n$.) In our testing problem, the tester is provided with $\rho^{\otimes n}$, which is invariant under any permutation of the n coordinates, and whether the tester accepts or rejects should be invariant under any unitary transformation of ρ. This means that if we measure $\rho^{\otimes n}$ in the *Schur basis* described in equation (6) below, we can throw away the information from the permutation and unitary registers without losing any relevant information. What is left is only the "irrep" label λ.

The end result is this: there is a sampling algorithm—referred to in [CHW07] as *weak Schur sampling*—which, on input a mixed state $\rho^{\otimes n}$, outputs a random partition $\boldsymbol{\lambda}$ whose distribution depends only on the spectrum of ρ. We will denote this distribution by SW_ρ^n. Furthermore, an argument which is essentially from [CHW07] (though see [MdW13, Lemma 19] for a full statement)

shows that for any spectrum property $\mathcal{P}$, there is an *optimal* tester in the model of Definition 1.2 whose operation is as follows: 1. Sample $\boldsymbol{\lambda} \sim \mathrm{SW}_\rho^n$. 2. Accept or reject based only on $\boldsymbol{\lambda}$. We may therefore proceed without loss of generality by analyzing only algorithms of this form. In particular, this means we need not study study quantum measurements or algorithms per se; in principle it suffices simply to understand the distribution SW_ρ^n (which is equivalent to the distribution on k-increasing subsequence lengths described in Fact 1.6).

In case ρ is the maximally mixed state, the distribution SW_ρ^n has been fairly well studied, starting with the works [TW01, Joh01, Bia01, Kup02] (see [Mél10a] for a recent, comprehensive treatment). It is known as the *Schur–Weyl* distribution, and we denote it by SW_d^n. (In the limit as $d \to \infty$, it approaches the well-known *Plancherel* distribution.) The exact distribution on partitions given by SW_d^n is somewhat complicated and difficult to work with, and so various works have instead sought to describe large-scale features of a "typical" $\boldsymbol{\lambda} \sim \mathrm{SW}_d^n$. For example, Biane [Bia01] showed that, up to small fluctuations, the "shape" of the random Young diagram $\boldsymbol{\lambda} \sim \mathrm{SW}_d^n$ tends toward a certain limiting shape Ω which depends only on the ratio $\frac{\sqrt{n}}{d}$. Furthermore, Meliot [Mél10a] has characterized these small fluctuations as being distributed according to a certain Gaussian process. The second of these results borrows heavily from a proof of the analogous result by Kerov (see [IO02]) for the Plancherel distribution, and we will give an overview his techniques below.

Kerov's approach involves studying a certain space of symmetric polynomial functions on Young diagrams. For example, if one is interested in showing that a random $\boldsymbol{\lambda} \sim \mathrm{SW}_d^n$ tends to have some coordinates which are much larger than the rest, then it would be natural to study "moments" of the form $\sum_i \boldsymbol{\lambda}_i^k$. However, the approach of Kerov would suggest studying the following "moments" instead:

$$p_k^*(\boldsymbol{\lambda}) \coloneqq \sum_{i=1}^{\infty} [(\boldsymbol{\lambda}_i - i + \tfrac{1}{2})^k - (-i + \tfrac{1}{2})^k], \text{ for } k \geq 1.$$

The polynomial family (p_k^*) inhabits (in fact, generates) the so-called *algebra of polynomial functions on the set of Young diagrams* Λ^* (also known as Kerov's *algebra of observables*). There are other important polynomial families within Λ^*—in addition to the p_k^* polynomials, our work involves the $\widetilde{p}_k$, c_k, $p_\mu^\sharp$, and s_μ^* polynomials—and each of these families sheds light on a different aspect of the input partition λ. For example, though the $p_\mu^\sharp(\lambda)$ polynomials don't give any obvious information regarding the "shape" of λ, they are unique in that we can easily compute the expectation $\mathbf{E}_{\boldsymbol{\lambda} \sim \mathrm{SW}_\rho^n}[p_\mu^\sharp(\boldsymbol{\lambda})]$ for any mixed state ρ. There exist some methods for passing from one polynomial family to another, and it is often the case that a problem most easily stated in terms of one polynomial family is most easily solved in terms of another.

The main component of our work is lower bounds for quantum spectrum testing, and these lower bounds generally have the following outline: 1. Reduce the problem to showing that a certain expression within the algebra of observables is small with high probability. 2. Use various polynomial-estimation techniques developed by Kerov and others for proving concentration of said expression. For example, roughly speaking the key component of the lower bound in Theorem 1.12 is showing that for $n \ll r^2$, the expression

$$\sum_{k=2}^{\infty} \frac{(-1)^k p_k^*(\boldsymbol{\lambda})}{k(r + \frac{1}{2})^k}$$

is typically very close to 0 when $\boldsymbol{\lambda} \sim \mathrm{SW}_r^n$. As another example, proving the lower bound in Theorem 1.10 reduces to showing that when $n \ll d/\epsilon^2$, the expression

$$\sum_{\substack{\text{partitions } \mu \text{ of } n \\ \text{with at most } d \text{ nonzeros}}} \frac{s_\mu^*(\boldsymbol{\lambda}) s_\mu(+2\epsilon, -2\epsilon, \ldots, +2\epsilon, -2\epsilon)}{\prod_{i=1}^d \prod_{j=1}^{\mu_i} (d + (j - i))}$$

is typically very close to 1 when $\boldsymbol{\lambda} \sim \mathrm{SW}_d^n$. Our upper bounds generally involve analyzing algorithms which accept or reject based on simple statistics of the sampled $\boldsymbol{\lambda} \sim \mathrm{SW}_d^n$. For example, the rank tester of Theorem 1.11 accepts if and only if the sampled $\boldsymbol{\lambda}$ has at most r nonzero parts, and the uniformity tester of Theorem 1.10 accepts if and only if the "content polynomial" $c_1(\boldsymbol{\lambda})$ is sufficiently small. As in the lower bounds, analyzing these algorithms uses techniques from the algebra of observables, and we sometimes also require certain combinatorial interpretations of the weak Schur sampling algorithm; e.g., its relationship with the Robinson–Schensted–Knuth "bumping" algorithm.

1.5 Acknowledgments

We thank Ilias Diakonikolas, Rocco Servedio, Greg Valiant, and Paul Valiant for helpful discussions regarding classical testing and learning of probability distributions. We thank Ashley Montanaro for helpful discussions regarding quantum property testing and for suggesting the proof of Proposition 2.2.

2 Preliminaries

2.1 Probabilistic distances

Given two discrete probability distributions $\mathcal{D}_1$ and $\mathcal{D}_2$ on a finite set Ω, the *total variation distance* between them is

$$d_{\mathrm{TV}}(\mathcal{D}_1, \mathcal{D}_2) \coloneqq \frac{1}{2} \cdot \sum_{\omega \in \Omega} |\mathcal{D}_1(\omega) - \mathcal{D}_2(\omega)| .$$

We will also require some nonsymmetric "distances" between probability distributions. The *chi-squared distance* is

$$d_{\chi^2}(\mathcal{D}_1, \mathcal{D}_2) \coloneqq \mathop{\mathbf{E}}_{\boldsymbol{\omega} \sim \mathcal{D}_2} \left[\left(\frac{\mathcal{D}_1(\boldsymbol{\omega})}{\mathcal{D}_2(\boldsymbol{\omega})} - 1 \right)^2 \right].$$

Further, if $\mathrm{supp}(\mathcal{D}_1) \subseteq \mathrm{supp}(\mathcal{D}_2)$, then the *Kullback–Leibler divergence* is

$$d_{\mathrm{KL}}(\mathcal{D}_1, \mathcal{D}_2) \coloneqq \mathop{\mathbf{E}}_{\boldsymbol{\omega} \sim \mathcal{D}_1} \left[\ln \left(\frac{\mathcal{D}_1(\boldsymbol{\omega})}{\mathcal{D}_2(\boldsymbol{\omega})} \right) \right].$$

To relate these quantities, Cauchy–Schwarz implies that $d_{\mathrm{TV}}(\mathcal{D}_1, \mathcal{D}_2) \leq \frac{1}{2}\sqrt{d_{\chi^2}(\mathcal{D}_1, \mathcal{D}_2)}$, and Pinsker's inequality states that $d_{\mathrm{TV}}(\mathcal{D}_1, \mathcal{D}_2) \leq \frac{1}{\sqrt{2}}\sqrt{d_{\mathrm{KL}}(\mathcal{D}_1, \mathcal{D}_2)}$.

We would also like to introduce a "permutation-invariant" notion of total variation distance. Suppose that the set Ω is naturally ordered; say, $\Omega = [d] \coloneqq \{1, 2, \ldots, d\}$. We define

$$d_{\mathrm{TV}}^{\mathrm{sym}}(\mathcal{D}_1, \mathcal{D}_2) \coloneqq d_{\mathrm{TV}}(\mathcal{D}_1^{\downarrow}, \mathcal{D}_2^{\downarrow}) = \min_{\pi \in \mathfrak{S}_d} \{d_{\mathrm{TV}}(\mathcal{D}_1, \mathcal{D}_2 \circ \pi)\}.$$

Here $\mathcal{D}_i^{\downarrow}$ denotes the probability distribution on $[d]$ given by rearranging $\mathcal{D}_i$'s probabilities in non-increasing order, so $\mathcal{D}_i^{\downarrow}(1) \geq \cdots \geq \mathcal{D}_i^{\downarrow}(d)$. By virtue of the permutation-invariance, we may also naturally extend this notation to the case when $\mathcal{D}_1$ and $\mathcal{D}_2$ are simply unordered multisets of nonnegative numbers summing to 1.

A d-dimensional mixed quantum state is represented as a *density matrix* $\rho \in \mathbb{C}^{d \times d}$; i.e., a positive semidefinite matrix with trace 1. We may write ρ using its spectral decomposition as

$$\rho = \sum_{i=1}^{d} \eta_i \cdot |\Psi_i\rangle\langle\Psi_i|,$$

where the $|\Psi_i\rangle$'s are orthornormal and the η_i's are nonnegative reals satisfying $\eta_1 + \cdots + \eta_d = 1$. Equivalently, ρ describes a probability distribution on pure states in which $|\Psi_i\rangle$ has probability η_i. If σ is another d-dimensional mixed state with eigenvalues $\{\lambda_1, \ldots, \lambda_d\}$ (thought of as a multiset), we will use the notation

$$d_{\mathrm{TV}}^{\mathrm{sym}}(\rho, \sigma) \coloneqq d_{\mathrm{TV}}^{\mathrm{sym}}(\{\eta_1, \ldots, \eta_d\}, \{\lambda_1, \ldots, \lambda_d\}).$$

We will now define *trace distance*, which is the standard notion of distance between two density matrices. (The above nonstandard notion of distance will be related to the trace distance in Proposition 2.2 below.) If $M \in \mathbb{C}^{d \times d}$ is any Hermitian matrix with eigenvalues $\eta_1, \ldots, \eta_d$, the *trace norm* of M is

$$\|M\|_{\mathrm{tr}} \coloneqq \mathrm{tr}\left(\sqrt{M^\dagger M}\right) = \sum_{i=1}^{d} |\eta_i|.$$

Given two density matrices ρ and σ, the *trace distance* between them is $d_{\mathrm{tr}}(\rho, \sigma) \coloneqq \frac{1}{2}\|\rho - \sigma\|_{\mathrm{tr}}$. The trace distance is the standard generalization of the total variation distance to mixed states; for example, it represents the maximum probability with which two mixed states can be distinguished by a measurement [NC10, equation (9.22)]. This property makes it the natural choice of distance for property testing of quantum states. We also have the following simple fact:

Fact 2.1. *Suppose ρ and σ are diagonal density matrices with diagonal entries $\eta = (\eta_1, \ldots, \eta_d)$ and $\lambda = (\lambda_1, \ldots, \lambda_d)$, respectively. Then $d_{\mathrm{tr}}(\rho, \sigma) = d_{\mathrm{TV}}(\eta, \lambda)$.*

2.2 Property testing

In the model of property testing, there is a set of objects $\mathcal{O}$ along with a distance measure $\mathrm{dist} : \mathcal{O} \times \mathcal{O} \to \mathbb{R}$. A property $\mathcal{P}$ is a subset of $\mathcal{O}$, and for an object $o \in \mathcal{O}$, we define the distance of o to $\mathcal{P}$ to be[2]

$$\mathrm{dist}(o, \mathcal{P}) \coloneqq \min_{o' \in \mathcal{P}} \{\mathrm{dist}(o, o')\}.$$

If $\mathrm{dist}(o, \mathcal{P}) \geq \epsilon$, then we say that o is ϵ-far from $\mathcal{P}$. A testing algorithm $\mathcal{T}$ tests $\mathcal{P}$ if, given some sort of "access" to $o \in \mathcal{O}$ (e.g., independent samples or queries), $\mathcal{T}$ accepts if $o \in \mathcal{P}$ and rejects if o is ϵ-far from $\mathcal{P}$. Generally, the aim is for $\mathcal{T}$ to be efficient according some measure, most typically the number of accesses made to o. (On the other hand, $\mathcal{T}$ is generally allowed unlimited computational power. Nevertheless, as we will see, all of the testers considered in this paper can be implemented efficiently.)

We will instantiate property testing in the following natural settings:

(i) **Properties of mixed states:** $\mathcal{O}$ is the set of d-dimensional mixed states ρ, the tester gets access to (unentangled) copies of ρ, and $\mathrm{dist} = d_{\mathrm{tr}}$.

(ii) **Unitarily invariant properties of mixed states:** As above, but $\mathcal{P}$ must be unitarily invariant; equivalently, whether or not $\rho \in \mathcal{P}$ only depends on the multiset of ρ's eigenvalues.

(iii) **Quantum spectrum testing:** $\mathcal{O}$ is the set of d-dimensional mixed states, $\mathcal{P}$ must be unitarily invariant, and $\mathrm{dist}(\rho, \sigma) = d_{\mathrm{TV}}^{\mathrm{sym}}(\rho, \sigma)$.

[2]Formally, our sets $\mathcal{O}$ will always lie within some $\mathbb{R}^N$ or $\mathbb{C}^N$, and we always require that $\mathcal{P}$ be a *closed* set. Thus the "min" here is well-defined.

(iv) **Symmetric properties of probability distributions:** $\mathcal{O}$ is the set of probability distributions $\mathcal{D}$ on $[d]$, the tester gets i.i.d. draws from $\mathcal{D}$, $\mathcal{P}$ is any symmetric property, and dist $= d_{\mathrm{TV}}$.

Let us now establish some basic facts about these models. The simplest fact is that Model (ii) is a special case of Model (i). Next, in Model (iv) it would be equivalent if we had chosen dist $= d_{\mathrm{TV}}^{\mathrm{sym}}$; this is by virtue of the assumption that $\mathcal{P}$ is a symmetric (permutation-invariant) property of distributions on $[d]$. Finally, we have the following important simplifying fact, whose proof is not trivial:

Proposition 2.2. *Models (ii) and (iii) are equivalent.*

Proof. We need to show that if $\mathcal{P}$ is a unitarily invariant property of d-dimensional mixed states then $d_{\mathrm{tr}}(\rho, \mathcal{P}) = d_{\mathrm{TV}}^{\mathrm{sym}}(\rho, \mathcal{P})$ holds for all mixed states ρ. By performing a unitary transformation, we may assume without loss of generality that ρ is a diagonal matrix with nonincreasing diagonal entries (spectrum).

The easy direction of the proof is showing that $d_{\mathrm{tr}}(\rho, \mathcal{P}) \leq d_{\mathrm{TV}}^{\mathrm{sym}}(\rho, \mathcal{P})$. To see this, suppose $\sigma \in \mathcal{P}$ achieves $d_{\mathrm{TV}}^{\mathrm{sym}}(\rho, \sigma) = \epsilon$. Let σ' denote the diagonal density matrix whose diagonal entries are the eigenvalues of σ arranged in nonincreasing order. Now σ' is unitarily equivalent to σ, and hence $\sigma' \in \mathcal{P}$ as well. But $d_{\mathrm{tr}}(\rho, \sigma') = \epsilon$ by Fact 2.1 and we therefore conclude $d_{\mathrm{tr}}(\rho, \mathcal{P}) \leq \epsilon$, as needed.

The more interesting direction is showing that $d_{\mathrm{TV}}^{\mathrm{sym}}(\rho, \mathcal{P}) \leq d_{\mathrm{tr}}(\rho, \mathcal{P})$. The authors learned the proof of this fact from Ashley Montanaro [Mon14]. Suppose that $\sigma \in \mathcal{P}$ achieves $d_{\mathrm{tr}}(\rho, \sigma) = \epsilon$. Since $\|\cdot\|_{\mathrm{tr}}$ is a unitarily invariant norm, a theorem of Mirsky (see [HJ13, Corollary 7.4.9.3]) states that

$$\|\rho - \sigma\|_{\mathrm{tr}} \geq \|\rho' - \sigma'\|_{\mathrm{tr}}, \tag{2}$$

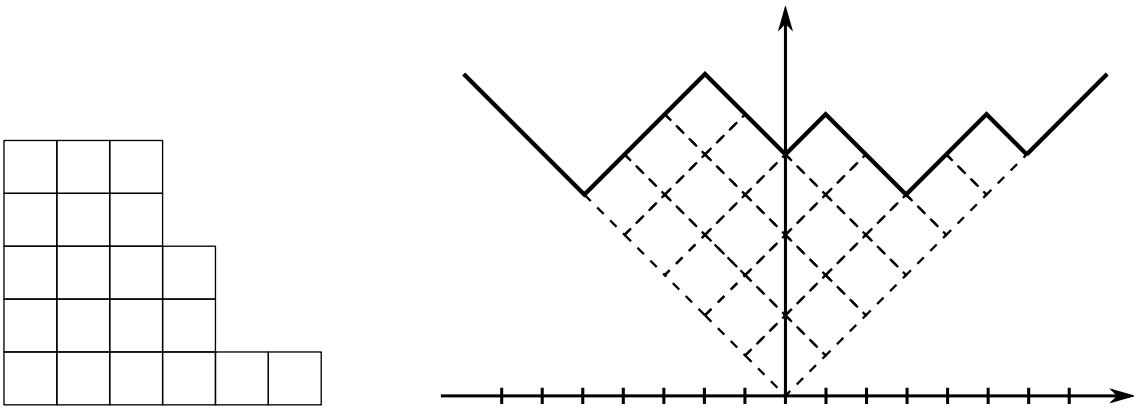
where σ' (respectively, ρ') denotes the diagonal density matrix whose entries are the eigenvalues of σ (respectively, ρ) arranged in nonincreasing order. We have $\rho' = \rho$, and σ' is again unitarily equivalent to σ, implying $\sigma' \in \mathcal{P}$. But the left-hand side of (2) is 2ϵ, and the right-hand side is $2d_{\mathrm{TV}}(\rho', \sigma')$ (by Fact 2.1), which in turn equals $2d_{\mathrm{TV}}^{\mathrm{sym}}(\rho, \sigma')$. Thus $d_{\mathrm{TV}}^{\mathrm{sym}}(\rho, \mathcal{P}) \leq \epsilon$, as needed. □

Finally, we remind the reader of Fact 1.7, which says that any quantum spectrum testing problem (in either of the equivalent Models (ii) and (iii)) is at least as hard as the corresponding classical problem in Model (iv).

2.3 Partitions and Young diagrams

A *partition* of $n \geq 1$, denoted $\lambda \vdash n$, is a list of nonnegative integers $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_k)$ satisfying $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k$ and $\lambda_1 + \lambda_2 + \ldots + \lambda_k = n$. The *length* of the partition, denoted $\ell(\lambda)$, is the number of nonzero λ_i's in λ. The partition's *size* is n, and is also written as $|\lambda|$. Two partitions are considered to be equivalent if they only differ in trailing zeros. For example, $(4, 2)$ and $(4, 2, 0, 0)$ are equivalent. We write Par to denote the set of all partitions, of any size. For $w \in \mathbb{N}^+$ we will use the notation $m_w(\lambda)$ to denote the number of parts i with $\lambda_i = w$. Finally, at one point we will require the fairly elementary fact (see e.g. [Rom14, (1.15)]) that the number of partitions of n is $2^{O(\sqrt{n})}$ (much more precise asymptotics are known [HR18]).

One way in which partitions arise is as *cycle types* of permutations $\pi \in \mathfrak{S}_n$. We say that π has cycle type $\lambda = (\lambda_1, \ldots, \lambda_k) \vdash n$ if π is the product of disjoint cycles of size $\lambda_1, \lambda_2, \ldots, \lambda_k$. (Note that π's length-1 cycles are included.) The standard notation for this is $\rho(\pi) = \lambda$. However we will use this notation extremely sparingly (and with warning) so as to preserve the symbol "ρ"



(a) French notation.

(b) Russian notation (in dashed lines). The marks on the horizontal axis are integral x-values, and the heavy black line is the curve $\lambda(x)$.

Figure 1: Two ways of drawing the partition $\lambda = (6, 4, 4, 3, 3)$.

for density matrices. In aid of this, we adopt the following convention: *whenever a permutation π appears in a place where a partition λ is expected, the meaning is that λ should be the cycle type of π*. We also use the following standard notation:

$$z_\lambda := \prod_{w \geq 1} (w^{m_w(\lambda)} \cdot m_w(\lambda)!).$$

When $\lambda \vdash n$, the quantity $n!/z_\lambda$ is the number of permutations in $\mathfrak{S}_n$ of cycle type λ, so z_λ^{-1} represents the probability that a uniformly random permutation in $\mathfrak{S}_n$ has cycle type λ.

It is standard to represent a partition $\lambda \vdash n$ pictorially with a *Young diagram*; i.e., a certain arrangement of n squares, called *cells* or *boxes*. There are several conventions for how to draw Young diagrams: we will define the *French notation*, the *Russian notation*, and the *Maya notation*. [3]

In the *French notation*, the Young diagram for $\lambda = (\lambda_1, \ldots, \lambda_k)$ is drawn with left-justified rows of cells: λ_1 cells in the bottom row, λ_2 cells on top of this, λ_3 cells on top of this, etc. As an example, the French notation for $(6, 4, 4, 3, 3)$ is pictured in Figure 1a. We think of the French diagram as consisting of unit squares sitting in $\mathbb{R}^2_+$, with bottom-left corner at the origin.

Given the French diagram, it's natural to define the *width* of λ as λ_1, and to refer to $\ell(\lambda)$ as its *height*. We can also define the *conjugate partition* of λ to be the partition $\lambda' \vdash n$ obtained by reflecting the French diagram through the line $y = x$; i.e., exchanging rows and columns. For example, the conjugate of $\lambda = (6, 4, 4, 3, 3)$ is $\lambda' = (5, 5, 5, 3, 1, 1)$. Note that the height of λ is the width of λ', and vice versa; in particular, we sometimes prefer the notation λ'_1 to $\ell(\lambda)$.

We now define the *Russian notation* for λ. This is obtained from the French notation by first rotating the diagram 45° counterclockwise about the origin, and then dilating by a factor of $\sqrt{2}$; see Figure 1b. The purpose of the dilation is so that the corners of the boxes will have integer x- and y-coordinates. The purpose of the rotation is so that conjugation corresponds to reflection in the y-axis and so that the boundary of the diagram forms the graph of a function:

[3] We will not require the *English notation*, which is the reflection of the French notation across the horizontal axis.

Definition 2.3. Given a partition λ drawn in Russian notation, its upper boundary forms the graph of a function with domain $[-\lambda'_1, \lambda_1] \subseteq \mathbb{R}$. We extend this function to have domain all of $\mathbb{R}$ according to the function $x \mapsto |x|$. We will use the notation $\lambda : \mathbb{R} \to \mathbb{R}_+$ for this function, which we remark is a continuous and piecewise linear curve. Any time we write $\lambda(x)$, where λ is a partition and $x \in \mathbb{R}$, we are referring to this curve. See Figure 1b for an example.

Finally, we define the *Maya notation*. It contains no boxes; just a sequence of black and white pebbles. However the Maya notation is typically drawn in conjunction with the Russian notation, with the pebbles being located on the half-integer points $\mathbb{Z} + \frac{1}{2}$ of the x-axis. In the Maya notation, a black pebble is placed at all points directly below a "downward-sloping" segment in λ's graph, and a white pebble is placed at all points directly below an "upward-sloping" segment. (Thus all sufficiently negative half-integer points have a black pebble and all sufficiently positive half-integer points have a white pebble.) The notation also includes a vertical tick mark to denote the location of the origin. A picture of the Russian and Maya notation for $\lambda = (6, 4, 4, 3, 3)$ appears later in Figure 4 (the reader consulting it now should ignore the red and green coloring, the dashed lines, and the box labeled "d"). One can check that the sequence of pebbles uniquely identifies the partition λ. It also uniquely determines the position of the origin mark, in that the number of black pebbles to the right of the origin mark always equals the number of white pebbles to the left of the origin mark. These numbers are both equal to $d(\lambda)$, defined to be the number of cells touching the y-axis in the Russian diagram. We make one more definition:

Definition 2.4. Given the Maya diagram of a partition λ, we may define its *modified Frobenius coordinates* to be the half-integer values $a_1^* > a_2^* > \cdots a_d^* > 0$ and $b_1^* > b_2^* > \cdots > b_d^* > 0$ (for $d = d(\lambda)$), where a_i^* is the position of the ith rightmost black pebble and b_i^* is the negative of the position of the ith leftmost white pebble. One may check that, equivalently, $a_i^* = \lambda_i - i + \frac{1}{2}$ and $b_i^* = \lambda'_i - i + \frac{1}{2}$. For example, if $\lambda = (6, 4, 4, 3, 3)$, then $a^* = (\frac{11}{2}, \frac{5}{2}, \frac{3}{2})$ and $b^* = (\frac{9}{2}, \frac{7}{2}, \frac{5}{2})$. The coordinates have the property that $\sum_i (a_i^* + b_i^*) = |\lambda|$.

For a partition λ (drawn either in the French or Russian notation), we often use the symbol "$\square$" to denote a box in λ's Young diagram. We write $[\lambda]$ for the set of all boxes in the diagram. Each box $\square \in [\lambda]$ is indexed by an ordered pair (i, j), where i is $\square$'s row and j is $\square$'s column. Note that this indexing is slightly peculiar vis-a-vis the French notation, in which the center of $\square$ has Cartesian coordinates $(j - \frac{1}{2}, i - \frac{1}{2})$. We define the *content* of cell $\square$ to be $c(\square) := j - i$. Note that in the Russian diagram, the content of $\square$ is the x-coordinate of its center. We also define the *hook length* $h(\square)$ of $\square$ via the French notation: it is the number of cells directly to the right or above $\square$, including $\square$ itself; equivalently, it is $(\lambda_i - j) + (\lambda'_j - i) + 1$.

Having defined "content" for cells in a Young diagram, we may introduce some convenient notation (essentially from [OO98b]) that generalizes the standard notions of "falling factorial power" and "rising factorial power". First, for $z \in \mathbb{R}$ and $m \in \mathbb{N}$, recall the *falling factorial power*[4]

$$z^{\downarrow m} := z(z-1)(z-2)\cdots(z-m+1)$$

and *rising factorial power*

$$z^{\uparrow m} := z(z+1)(z+2)\cdots(z+m-1).$$

We generalize this notation to the case of an arbitrary partition $\lambda \vdash m$:

$$z^{\downarrow \lambda} := \prod_{\square \in [\lambda]} (z - c(\square)) \quad \text{and} \quad z^{\uparrow \lambda} := \prod_{\square \in [\lambda]} (z + c(\square)).$$

[4] Or Pochhammer symbol, sometimes denoted $(z)_m$ or $z^{\underline{m}}$.

6	7	8			
5	6	6			
3	3	4	5		
2	2	3	3		
1	1	1	2	3	3

Figure 2: A semistandard tableau of shape $\lambda = (6, 4, 4, 3, 3)$ with alphabet $[8]$.

2.3.1 Random words and Young diagrams, and symmetric polynomials

Definition 2.5. Let $\mathcal{A}$ be an *alphabet*; i.e., a totally ordered set. Most often we consider $\mathcal{A} = [d]$. A *word* is a finite sequence $(a_1, \ldots, a_n)$ of elements from $\mathcal{A}$. It is *weakly increasing* if $a_1 \leq a_2 \leq \cdots \leq a_n$ and *strongly (or strictly) increasing* if $a_1 < a_2 < \cdots < a_n$. If $\mathcal{D}$ is a probability distribution on $\mathcal{A}$ we write $\mathcal{D}^{\otimes n}$ to denote the probability distribution on words of length n given by drawing the letters independently from $\mathcal{D}$.

Definition 2.6. Given a word $a \in [d]^n$, there is an associated partition $\lambda \vdash n$ of length at most d called the *sorted type (or histogram)*. It is defined as follows: λ_i is the frequency of the ith-most frequent letter in a, for $1 \leq i \leq d$. In other words, λ is the histogram of letter frequencies, sorted into nonincreasing order. For example, the sorted type of $(4, 1, 3, 4, 4, 4, 1, 4) \in [4]^8$ is $(5, 2, 1, 0) \vdash 8$.

Definition 2.7. Let $x_1, \ldots, x_d$ be indeterminates, typically standing for real numbers. For $m \in \mathbb{N}$, the *mth power sum symmetric polynomial* is $p_m(x) = \sum_{i=1}^{d} x_i^m$. More generally, for a partition λ we define $p_\lambda(x) = \prod_{i=1}^{\ell(\lambda)} p_{\lambda_i}(x)$. By our conventions, if $\pi \in \mathfrak{S}_n$ then $p_\pi(x)$ denotes $p_\lambda(x)$, where λ is the cycle type of π. If $\mathcal{D} = (\eta_1, \ldots, \eta_d)$ is a probability distribution on $[d]$, there is a natural interpretation of $p_\pi(\eta_d, \ldots, \eta_d)$: it is the probability that a random word $\boldsymbol{a} \sim \mathcal{D}^{\otimes n}$ is invariant under the permutation π.

Definition 2.8. Let $\lambda \vdash n$, and think of its Young diagram in the French notation. If each cell is filled with an element from some alphabet $\mathcal{A}$, we call the result a *Young tableau of shape* λ. The Young tableau is said to be *semistandard* if its entries are weakly increasing from left-to-right along rows and are strongly increasing from bottom-to-top along columns. Figure 2 gives an example semistandard tableau of shape $(6, 4, 4, 3, 3)$. If the rows are in fact strongly increasing, the Young tableau is called *standard*.

Definition 2.9. For reasons we will see later, the number of standard Young tableaus[5] of shape $\lambda \vdash n$ over alphabet $[n]$ is denoted $\dim(\lambda)$. It can be computed via the *Hook-Length Formula* of Frame, Robinson, and Thrall [FRT54] (see also [Sta99, Corollary 7.21.6]):

$$\dim(\lambda) = \frac{n!}{\prod_{\square \in [\lambda]} h(\square)}.$$

We will also consider counting semistandard tableaus, via the following definition:

[5] Often spelled "tableaux".

Definition 2.10. Let $x_1, \ldots, x_d$ be indeterminates, typically standing for real numbers. Given $\lambda \vdash n$, the *Schur polynomial* $s_\lambda(x_1, \ldots, x_d)$ is the degree-n homogeneous polynomial defined by $\sum_T x^T$, where the sum is over all semistandard tableaus of shape λ over alphabet $[d]$, and where

$$x^T := \prod_{i=1}^{d} x_i^{(\#\text{ of occurrences of letter } i \text{ in } T)}.$$

The following formula from [Sta99, Corollary 7.21.4] thereby lets us count the number of such tableaus:

$$s_\lambda(\underbrace{1, 1, \ldots, 1}_{d \text{ entries}}) = \frac{d^{\uparrow\lambda}}{\prod_{\square \in [\lambda]} h(\square)}.$$

We record here a consequence of the above two formulas:

Proposition 2.11. *Let λ be a partition and let $d \in \mathbb{Z}^+$. Then* $s_\lambda(\underbrace{1, \ldots, 1}_{d \text{ entries}}) = \dfrac{(\dim \lambda) d^{\uparrow\lambda}}{|\lambda|!}$.

When $\ell(\lambda) > d$, there are no semistandard tableaus of shape λ over alphabet $[d]$. Thus, the sum $\sum_T x^T$ is the empty sum. This gives us the following fact about Schur polynomials:

Proposition 2.12. *Consider the Schur polynomial $s_\lambda(x_1, \ldots, x_d)$. If $\ell(\lambda) > d$ then $s_\lambda \equiv 0$.*

Though it is not at all obvious from the definition, the Schur polynomials are symmetric. This can be inferred from the following classical fact (see e.g. [Sta99, Theorem 7.15.1]), which expresses them as the ratio of a skew-symmetric polynomial and the Vandermonde determinant:

Theorem 2.13. $s_\lambda(x_1, \ldots, x_d) = \dfrac{\det\left(x_i^{d+\lambda_j - j}\right)_{ij}}{\det\left(x_i^{d-j}\right)_{ij}}$.

We will actually not need this formula. Instead, we will next describe a combinatorial algorithm which gives an interpretation for $s_\lambda(\eta_1, \ldots, \eta_d)$ when $\mathcal{D} = (\eta_1, \ldots, \eta_d)$ is a probability distribution.

2.3.2 The RSK algorithm

We now describe the *Robinson–Schensted–Knuth (RSK)* algorithm $\mathrm{RSK}(\cdot)$, which takes as input a word $a \in \mathcal{A}^n$ and outputs a partition $\lambda = \mathrm{RSK}(a) \vdash n$. The relevance of RSK to quantum spectrum testing is described at the end of this section. As there are many descriptions of the RSK algorithm in the literature (see, e.g., [Knu70, Bay02, Dor05, Rom14]), we will be brief.

The RSK algorithm. Given as input a word $a = (a_1, \ldots, a_n)$ over (ordered) alphabet $\mathcal{A}$, the RSK algorithm produces a sequence $T_0, \ldots, T_n$ of semistandard tableaus over $\mathcal{A}$, with T_i having size i (and being thought of in French notation). Tableau T_{i+1} is produced from tableau T_i via the "insertion" of letter a_i into the 1st row. The insertion algorithm for letter b into row j of tableau T is as follows: Find the rightmost position in the jth row such that if b were placed there, weak-increasingness along row j would be maintained. If this position is at the end of the row, the insertion of b is complete. If instead it is at a cell that already contains some letter c (which will in fact be the least c in row j with $c > b$) then c is "bumped up". By this we mean that the insertion algorithm is recursively applied to letter c and row $j + 1$ of T (which may be a newly created row, in which the insertion will immediately terminate with c in its own row at the top of T). In the end, the output of the RSK algorithm is the Young diagram $\lambda \vdash n$ given by the *shape* of T_n; i.e., $\mathrm{RSK}(a)$ is T_n with its cell entries erased.

To get some feel for this algorithm, note that if the inserted word a is weakly increasing then $\mathrm{RSK}(a) = (n) \vdash n$. On the other hand, if a is strongly decreasing, the output will be $\mathrm{RSK}(a) = (1, 1, \ldots, 1) \vdash n$. More generally, it is not hard to show that when $\mathrm{RSK}(a) = \lambda$, the value λ_1 is the length of the longest weakly increasing subsequence of a, and $\ell(\lambda) = \lambda'_1$ is the length of the longest strongly decreasing subsequence of a. Even more generally, we have the following theorem of Greene [Gre74], completely characterizing the partition $\mathrm{RSK}(a)$ in terms of increasing subsequences:

Theorem 2.14. *Let* $\mathrm{RSK}(a) = \lambda$. *Then for each* $k \geq 1$, *the value* $\lambda_1 + \ldots + \lambda_k$ *is the length of the longest* k*-increasing subsequence in* a *(as defined in Fact 1.6).*

Indeed, the RSK algorithm is most commonly used in the literature to study the length of the longest increasing subsequence of a random permutation (equivalently, of a random word $\boldsymbol{a} \sim \mathcal{X}^{\otimes n}$, where $\mathcal{X}$ denotes the uniform distribution on the alphabet $\mathcal{A} = [0, 1]$).

Let us note one immediate consequence of Greene's theorem. (This consequence may also be derived directly from the description of the RSK algorithm.)

Proposition 2.15. *Given* $a \in [d]^n$, *let* $\mathrm{RSK}(a) = \lambda$. *Write* $c_i(a)$ *for the number of letter* i*'s in* a. *Then* λ *majorizes* $c(a) := (c_1(a), \ldots, c_d(a))$.

To see why this is true, note that for each $k \in [d]$, the all one's, all two's, ..., and all k's subsequences together form a k-increasing subsequence of size $c_1(a) + \ldots + c_k(a)$, which by Theorem 2.14 is at most $\lambda_1 + \ldots + \lambda_k$, giving the proposition. As $c(a)$ is the histogram of a, this shows that we can view $\mathrm{RSK}(a)$ as a "shifted histogram" of a in which cells are shifted towards the lower numbers.

Although Greene's Theorem succinctly characterizes the output by the RSK algorithm, it is important to retain the algorithm itself and even to consider an extension of it. Suppose that when the RSK algorithm is applied to a we also form a standard tableau T' over alphabet $[n]$, where T' has the same shape as T_n and each cell $\square$ in T' is labeled by the "time" at which $\square$ was created in T_n. As noted by Knuth [Knu70], the word a is uniquely determined by the pair (T_n, T'). As a consequence of this and of previous formulas, it is not hard to verify the following important fact, perhaps first observed by Its, Tracy, and Widom [ITW01, equation (2-1)]:

Proposition 2.16. *Let* $\boldsymbol{a} \sim \mathcal{D}^{\otimes n}$, *where* $\mathcal{D} = (\eta_1, \ldots, \eta_d)$ *is a probability distribution on* $[d]$. *Then for each* $\lambda \vdash n$,

$$\mathbf{Pr}[\mathrm{RSK}(\boldsymbol{a}) = \lambda] = \dim(\lambda) \cdot s_\lambda(\eta_1, \ldots, \eta_d).$$

By the symmetry of the Schur polynomials, this implies the surprising fact that the distribution of $\mathrm{RSK}(\boldsymbol{a})$ is invariant to permutations of $\mathcal{D}$.

Finally, we mention the connection between the RSK algorithm and quantum spectrum testing. As we will eventually see in Section 2.6 (Remark 2.24), all of quantum spectrum testing can be boiled down to classical testing of symmetric probability distributions $\mathcal{D}$, with the following twist: Rather than getting to see a random word $\boldsymbol{a}$ sampled from $\mathcal{D}^{\otimes n}$, the tester only gets to see the partition $\boldsymbol{\lambda} = \mathrm{RSK}(\boldsymbol{a})$. In light of Greene's Theorem 2.14, this statement is equivalent to Fact 1.6.

2.4 Representation theory, and the symmetric group

Herein we recall some basics of representation theory. We will mainly focus on $\mathbb{C}$-representations of finite groups G (though at one point we will want to consider representations of the unitary group). We may therefore define a *representation* μ of G to be a group homomorphism from G into U_d, for some $d \in \mathbb{Z}^+$. Here U_d denotes the group of $d \times d$ unitary matrices. The number d is also called the *dimension* of the representation μ and is denoted $\dim(\mu)$.

Two representations μ_1 and μ_2 are said to be *isomorphic* if there is some unitary matrix U such that $U\mu_1 U^\dagger = \mu_2$. In this case we write $\mu_1 \cong \mu_2$. The *direct sum* of k representations $\mu_1, \ldots, \mu_k$ produces the representation μ given by block-diagonal matrices:

$$\mu(g) \coloneqq \begin{bmatrix} \mu_1(g) & 0 & \ldots & 0 \\ 0 & \mu_2(g) & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \mu_k(g) \end{bmatrix} \tag{3}$$

for all $g \in G$. Equivalently, we may write

$$\mu(g) \coloneqq \sum_{i=1}^{k} |i\rangle\langle i| \otimes \mu_i(g). \tag{4}$$

We will also write $\mu = \mu_1 \oplus \ldots \oplus \mu_k$ to denote that μ is the direct sum of $\mu_1, \ldots, \mu_k$.

Let μ_1 be a representation of the group G_1 and μ_2 be a representation of the group G_2. Then the *tensor product* of μ_1 and μ_2, denoted $\mu_1 \otimes \mu_2$, is the representation defined by

$$(\mu_1 \otimes \mu_2)(g, h) \coloneqq (\mu_1(g)) \otimes (\mu_2(h)),$$

where the right-hand side uses the ordinary matrix tensor product. We have $\dim(\mu_1 \otimes \mu_2) = \dim(\mu_1) \cdot \dim(\mu_2)$.

In our setting, a representation μ of G is said to be *reducible* if there are representations μ_1 and μ_2 such that $\mu \cong \mu_1 \oplus \mu_2$. Otherwise it is *irreducible*, and is often called an *irrep* for brevity. Every representation can be uniquely decomposed into a direct sum of irreps (up to isomorphism and rearrangement of summands). Further, the set of all irreps of G (up to isomorphism), denoted $\widehat{G}$, is finite. Indeed, if we define the *regular representation* of G to be the $|G|$-dimensional representation R given by $R(g) = \sum_{h \in G} |gh\rangle\langle h|$, then R's decomposition into irreps contains every $\mu \in \widehat{G}$, with μ occurring $\dim(\mu)$ times. As a consequence, we have the formula

$$|G| = \sum_{\mu \in \widehat{G}} (\dim \mu)^2.$$

This fact leads to a natural *probability distribution* on irreps of G:

Definition 2.17. For a finite group G, the *Plancherel distribution* is the probability distribution on irreps in which $\mu \in \widehat{G}$ has probability $(\dim \mu)^2/|G|$.

For a group G and a representation μ, the character χ_μ is the function $\chi_\mu : G \to \mathbb{C}$ defined by

$$\chi_\mu(g) = \operatorname{tr}(\mu(g)),$$

for each $g \in G$. We have the following simple fact:

Fact 2.18. *Let μ be a representation of G. Then χ_μ is a* class function*; i.e., it is constant on the conjugacy classes of G.*

We now recall some basics of Fourier analysis over an arbitrary finite group G (though we will ultimately only need the case $G = \mathfrak{S}_n$). For $f, g : G \to \mathbb{C}$ we define $\langle f, g \rangle = \mathbf{E}_{\boldsymbol{u} \sim G}[f(\boldsymbol{u})\overline{g(\boldsymbol{u})}]$. Under this inner product, the characters $(\chi_\mu)_{\mu \in \widehat{G}}$ form an orthonormal basis for the space of class functions $f : G \to \mathbb{C}$. For general $f, g : G \to \mathbb{C}$ we define $(f * g)(u) = \mathbf{E}_{\boldsymbol{v} \sim G}[f(\boldsymbol{v})g(\boldsymbol{v}^{-1}u)]$; this

includes a nonstandard normalization by $\frac{1}{|G|}$. For a class function f and $\mu \in \widehat{G}$ we employ the following "Fourier notation": $\widetilde{f}(\mu) = \langle f, \chi_\mu \rangle$. (According to standard notation we would have $\widetilde{f}(\mu) = \frac{1}{|G|}\mathrm{tr}\left(\widehat{f}\right)$). Then Fourier inversion is simply $f = \sum_\mu \widetilde{f}(\mu)\chi_\mu$. Further, if g is another class function we have the formula $\widetilde{f * g}(\mu) = \frac{1}{\dim \mu}\widetilde{f}(\mu)\widetilde{g}(\mu)$.

We close this section by specifically discussing the representation theory of the symmetric group $\mathfrak{S}_n$. Two permutations $\pi, \sigma \in \mathfrak{S}_n$ are conjugate within the group $\mathfrak{S}_n$ if and only if they have the same cycle type. As a result, the conjugacy classes of $\mathfrak{S}_n$ can be identified with the partitions of n. As it happens, the set $\widehat{\mathfrak{S}}_n$ of irreps of the symmetric group can *also* be naturally identified with the partitions of n. For $\lambda \vdash n$, we will use the notation $\mathtt{p}_\lambda$ for the corresponding irrep of $\mathfrak{S}_n$. (To avoid getting too far afield, we will not actually describe the representation $\mathtt{p}_\lambda$.) Recalling Fact 2.18, we introduce the following notation:

Definition 2.19. Let $\lambda \vdash n$. We denote the character $\chi_{\mathtt{p}_\lambda}$ more simply as χ_λ. We remark that χ_λ is known to take on only rational values; in particular, $\overline{\chi_\lambda} = \chi_\lambda$. If $\mu \vdash n$ then we let $\chi_\lambda(\mu)$ denote $\chi_\lambda(\pi)$, where $\pi \in \mathfrak{S}_n$ is any permutation with cycle type μ. This is well defined since χ_λ is constant on the conjugacy classes of $\mathfrak{S}_n$. Finally, we also write $\dim(\lambda)$ for $\dim(\mathtt{p}_\lambda)$. It is well known [Sag01, Theorem 2.6.5] that $\dim(\lambda)$ is equal to the number of standard Young tableaus of shape λ over alphabet $[n]$, explaining the notation from Definition 2.9.

Following Stanley [Sta99, Corollary 7.17.5], we can actually give a definition of the symmetric group characters χ_μ in terms of the power sum and Schur polynomials:

Theorem 2.20. *In the context of Fourier analysis over the group $G = \mathfrak{S}_n$, suppose $\mu \vdash n$ and $x \in \mathbb{C}^d$. Then $p_{(\cdot)}(x) \coloneqq \pi \mapsto p_\pi(x)$ is a class function, and its Fourier coefficients are given by*

$$\widetilde{p_{(\cdot)}(x)}(\mu) = s_\mu(x).$$

Although this can be taken as an implicit definition of the characters χ_μ, we will more often think of the characters χ_μ as "known" and of Theorem 2.20 as letting us express the Schur polynomials in terms of the power sum polynomials.

2.5 Weak Schur sampling

In this section we will introduce the weak Schur sampling algorithm. Our treatment of this topic will heavily follow the treatments given in Aram Harrow's thesis [Har05] and the paper [CHW07].

To motivate the algorithm let us briefly consider the classical problem of testing symmetric properties of probability distributions on $[d]$. In this model, the tester obtains a random word $\boldsymbol{a} = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n)$, where each letter $\boldsymbol{a}_i$ is drawn independently from an unknown distribution $\mathcal{D}$ on $[d]$. The tester wants to decide whether $\mathcal{D}$ satisfies a certain symmetric property $\mathcal{P}$. Since the samples $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ are independent, the tester could—without loss of generality—randomly permute them according to any $\pi \in \mathfrak{S}_n$. Similarly, since the property $\mathcal{P}$ is symmetric, the tester could—again, without loss of generality—simultaneously apply any permutation $\sigma \in \mathfrak{S}_d$ to the letters it sees. Roughly speaking, the tester can "factor out" the action of the group $\mathfrak{S}_n \times \mathfrak{S}_d$. The information that remains is precisely the sorted type $\boldsymbol{\lambda} \vdash n$ of $\boldsymbol{a}$ (recall Definition 2.6).[6] Thus we see that the task of analyzing property testing of symmetric probability distributions boils down to the task of understanding the random partition $\boldsymbol{\lambda} \vdash n$ (of length at most d) induced as the sorted type of a random word drawn from $\mathcal{D}^{\otimes n}$.

[6] This partition carries the same information as the so-called "fingerprint" used in classical property literature [Bat01, Val08].

A similar but more complicated state of affairs holds for quantum spectrum testing. In this case, there is an unknown d-dimensional mixed state ρ, and the tester may measure n copies, $\rho^{\otimes n}$, in an attempt to determine whether ρ satisfies a certain unitarily-invariant property $\mathcal{P}$. As before, the tester could (without loss of generality) randomly permute the copies according to any $\pi \in \mathfrak{S}_n$. And in this quantum scenario, by the unitary-invariance of $\mathcal{P}$, the tester could also (without loss of generality) simultaneously apply any unitary $U \in U_d$ to each copy. *Weak Schur sampling* refers to the process of "factoring out" this action of $\mathfrak{S}_n \times U_d$. What remains is again a random partition $\boldsymbol{\lambda} \vdash n$ of length at most d, whose distribution depends only on the spectrum of ρ. (In fact, as we will see later in Remark 2.24, the distribution of $\boldsymbol{\lambda}$ is precisely that of $\mathrm{RSK}(\boldsymbol{a})$ where $\boldsymbol{a}$ is a random word chosen according to the probability distribution on $[d]$ defined by ρ's spectrum.) To understand this situation more thoroughly, we will need to discuss representation theory in more detail.[7]

As mentioned above, the groups $\mathfrak{S}_n$ and U_d each have a natural, unitary action on the space $(\mathbb{C}^d)^{\otimes n}$; the associated representations $\mathtt{P}$ and $\mathtt{Q}$ (respectively) are defined on the standard basis vectors $|a_1\rangle \otimes |a_2\rangle \otimes \cdots \otimes |a_n\rangle$ (for $a_i \in [d]$) via

$$\begin{aligned}\mathtt{P}(\pi)\ |a_1\rangle \otimes |a_2\rangle \otimes \ldots \otimes |a_n\rangle &= |a_{\pi^{-1}(1)}\rangle \otimes |a_{\pi^{-1}(2)}\rangle \otimes \ldots \otimes |a_{\pi^{-1}(n)}\rangle,\\ \mathtt{Q}(U)\ |a_1\rangle \otimes |a_2\rangle \otimes \ldots \otimes |a_n\rangle &= \ (U|a_1\rangle) \otimes (U|a_2\rangle) \otimes \ldots \otimes (U|a_n\rangle).\end{aligned}$$

We know the irreps of $\mathfrak{S}_n$ are indexed by partitions of n; thus, the representation $\mathtt{P}$ must decompose as

$$\mathtt{P}(\pi) \overset{\mathfrak{S}_n}{\cong} \bigoplus_{\lambda \vdash n} \mathtt{p}_\lambda(\pi) \otimes I_{m_\lambda}, \tag{5}$$

with m_λ denoting the number of copies of $\mathtt{p}_\lambda$ in the decomposition. The representation $\mathtt{Q}$ also decomposes into irreps of the group U_d. As it happens, these (infinitely many) irreps can *also* be naturally identified with partitions; specifically, for each partition $\lambda \in \mathrm{Par}$ with length at most d, there is an associated irrep $\mathtt{q}^d_\lambda \in \widehat{U}_d$. Furthermore, the theory of *Schur–Weyl duality* states that there is significant joint structure to these two decompositions. This structure ultimately arises because the two representations $\mathtt{P}$ and $\mathtt{Q}$ commute (i.e., $\mathtt{P}(\pi)\mathtt{Q}(U) = \mathtt{Q}(U)\mathtt{P}(\pi)$ for all $\pi \in \mathfrak{S}_n$, $U \in U_d$), and hence the simultaneous action $\mathtt{PQ}$ defined by $\mathtt{PQ}(\pi, U) \coloneqq \mathtt{P}(\pi)\mathtt{Q}(U)$ is a representation of the direct product group $\mathfrak{S}_k \times U_d$.

Schur–Weyl duality. $\mathtt{PQ} \overset{\mathfrak{S}_n \times U_d}{\cong} \bigoplus_{\lambda \vdash n} \mathtt{p}_\lambda \otimes \mathtt{q}^d_\lambda$.

In particular, by taking $U = id$ we see that m_λ, the multiplicity of $\mathtt{p}_\lambda$ in the decomposition of $\mathtt{P}$, is equal to $\dim(\mathtt{q}^d_\lambda)$. Similarly, by taking $\pi = id$, we see that the multiplicity of $\mathtt{q}^d_\lambda$ in the decomposition of $\mathtt{Q}$ is $\dim(\lambda) = \dim(\mathtt{p}_\lambda)$.

To restate Schur–Weyl duality, there exists a certain $d^n \times d^n$ unitary matrix U_{Schur} such that

$$U_{\mathrm{Schur}}\mathtt{P}(\pi)\mathtt{Q}(U)U^\dagger_{\mathrm{Schur}} = \sum_{\lambda \vdash n} |\lambda\rangle\langle\lambda| \otimes \mathtt{p}_\lambda(\pi) \otimes \mathtt{q}^d_\lambda(U), \tag{6}$$

for all $\pi \in \mathfrak{S}_n$, $U \in U_d$. We view U_{Schur} as a unitary linear transformation that performs a change of basis, from the standard basis into the *Schur basis*. We may now state the weak Schur sampling algorithm:

[7] In particular, we will go slightly beyond the framework from Section 2.4 by mentioning representations of the unitary group, which is of course not a finite group.

Weak Schur sampling. Given $\rho^{\otimes n}$, where ρ is a d-dimensional mixed state, the weak Schur sampling algorithm works as follows:

1. Measure $\rho^{\otimes n}$ in the Schur basis, receiving basis state $|\boldsymbol{\lambda}\rangle \otimes |\boldsymbol{p}\rangle \otimes |\boldsymbol{q}\rangle$.
2. Output $\boldsymbol{\lambda}$, a partition of size n and length at most d.

We will write SW_ρ^n for the distribution on partitions induced from $\rho^{\otimes n}$ by the weak Schur sampling algorithm. We will also use the shorthand

$$\mathrm{SW}_\rho^n(\lambda) \coloneqq \Pr_{\boldsymbol{\lambda} \sim \mathrm{SW}_\rho^n}[\boldsymbol{\lambda} = \lambda].$$

As we will state shortly, performing weak Schur sampling is without loss of generality in the context of testing unitarily invariant properties. To see why, suppose ρ is a d-dimensional mixed state, and consider the product mixed state $\rho^{\otimes n}$. Then it's not too hard to show (using invariance under $\mathtt{P}$ and Schur's Lemma, see e.g. [Har05, equation (6.1)]) that when represented in the Schur basis, it has a "trivial $\mathfrak{S}_n$ register":

Fact 2.21. *We may write* $U_{\mathrm{Schur}} \rho^{\otimes n} U_{\mathrm{Schur}}^\dagger = \sum_{\lambda \vdash n} |\lambda\rangle\langle\lambda| \otimes I \otimes R_\lambda^\rho$, *for some matrices* R_λ^ρ. *Here, for each* λ *we interpret* I *as the* $\dim(\lambda) \times \dim(\lambda)$ *identity matrix.*

As a consequence, it makes sense that a testing algorithm may discard the $\mathfrak{S}_n$ register. Now in general, the "U_d register" R_λ^ρ of $\rho^{\otimes n}$ is not trivial, and thus it may seem like the tester is losing information by discarding it. (Indeed, this potential loss is the source of the word "weak" in the phrase "weak Schur sampling".) However when testing unitarily invariant properties of ρ, the state $\rho^{\otimes n}$ should be treated no differently than the state $\mathtt{Q}(U)\rho^{\otimes n}\mathtt{Q}(U^\dagger) = (U\rho U^\dagger)^{\otimes n}$, for any $U \in U_d$. In particular, a tester could average over all unitaries U, and this *would* cause the resulting state to have trivial a U_d register in the Schur basis. This idea is formalized in the next lemma, which shows that weak Schur sampling is an optimal quantum measurement for the testing of unitarily invariant properties. The lemma, implicit in [CHW07], can be found with proof in [MdW13, Lemma 19].

Lemma 2.22. *Let* $\mathcal{P}$ *be a unitarily invariant property of* d*-dimensional mixed states. Assume there exists a tester which uses* n *copies of the input state* ρ*, accepts all states* $\rho \in \mathcal{P}$ *with probability at least* $1 - \delta$*, but accepts all states which are* ϵ*-far from* $\mathcal{P}$ *with probability at most* $1 - f(\epsilon)$ *for* $\epsilon > 0$*. Then there exists a tester with the same parameters which consists of performing weak Schur sampling on* $\rho^{\otimes n}$ *and then classically postprocessing the results.*

As a result of this lemma, we are able to focus exclusively on the weak Schur sampling algorithm in this paper. One final remark: Although our quantum spectrum testing upper bounds are formally only concerned with copy complexity, they can in fact also be implemented *efficiently*, by (quantum) algorithms running in time $\mathrm{poly}(n, \log d, \log(1/\epsilon))$. This holds because the only expensive operation is the computation of the Schur change-of-basis, and this can be done in $\mathrm{poly}(n, \log d, \log(1/\epsilon))$ time; see [BCH05, Appendix A], [Har05, Section 8.1.1].

2.6 Understanding the weak Schur sampling distribution

There are a several ways to understand the probability distribution induced by weak Schur sampling algorithm, each of which proves advantageous in different settings. Let us begin with a direct calculation that expresses the probabilities in terms of the Schur polynomials. The following known fact may be attributed to Alicki et al. [ARS88]; see [Aud06, equation (36)] for further discussion. We will include a proof for the reader's convenience.

Proposition 2.23. *Let ρ be a d-dimensional density matrix with eigenvalues $\eta_1, \eta_2, \ldots, \eta_d$. Then*

$$\mathrm{SW}^n_\rho(\lambda) = \dim(\lambda) \cdot s_\lambda(\eta_1, \eta_2, \ldots, \eta_d).$$

In particular, SW^n_ρ depends only on the spectrum of ρ.

Remark 2.24. As this is the exact same formula as in Proposition 2.16, we conclude that if $\mathcal{D}$ is the probability distribution on $[d]$ given by the spectrum of ρ (in any order), then

$$\mathrm{SW}^n_\rho(\lambda) = \mathop{\mathbf{Pr}}_{\boldsymbol{a} \sim \mathcal{D}^{\otimes n}}[\mathrm{RSK}(\boldsymbol{a}) = \lambda].$$

This gives a completely "quantum-free" way of analyzing quantum spectrum testing, as mentioned in Fact 1.6. Nevertheless, we will actually use this fact only occasionally (mainly via Theorem 2.14). As we will see later, interpreting SW^n_ρ via representation theory proves to be more powerful.

Proof of Proposition 2.23. By definition, $\mathrm{SW}^n_\rho(\lambda) = \mathrm{tr}(\Pi_\lambda \rho^{\otimes n})$, where Π_λ denotes the operator that projects onto the subspace corresponding to λ in the Schur basis. It is a basic fact of representation theory (following from orthogonality relations, see e.g. [CHW07, Equation (7)]) that from the decomposition (5) of $\mathtt{P}$ we may deduce

$$\Pi_\lambda = \dim(\lambda) \cdot \mathop{\mathbf{E}}_{\boldsymbol{\pi} \sim \mathfrak{S}_n} \left[\overline{\chi_{\mathtt{p}_\lambda}(\boldsymbol{\pi})} \cdot \mathtt{P}(\boldsymbol{\pi}) \right] = \dim(\lambda) \cdot \mathop{\mathbf{E}}_{\boldsymbol{\pi} \sim \mathfrak{S}_n} \left[\chi_\lambda(\boldsymbol{\pi}) \cdot \mathtt{P}(\boldsymbol{\pi}) \right].$$

Thus

$$\mathrm{SW}^n_\rho(\lambda) = \dim(\lambda) \cdot \mathop{\mathbf{E}}_{\boldsymbol{\pi} \sim \mathfrak{S}_n} \left[\chi_\lambda(\boldsymbol{\pi}) \cdot \mathrm{tr}(\mathtt{P}(\boldsymbol{\pi}) \rho^{\otimes n}) \right].$$

To compute the trace, we may assume by unitary invariance that $\rho = \mathrm{diag}(\eta_1, \ldots, \eta_d)$. Thus

$$\rho^{\otimes n} = \sum_{\substack{\text{words} \\ (a_1, \ldots, a_n) \in [d]^n}} \left(\prod_{i=1}^n \eta_{a_i} \right) |a_1, \ldots, a_n\rangle \langle a_1, \ldots, a_n|.$$

Notice that if we let $\mathcal{D}_\eta$ denote the probability distribution on $[d]$ in which $\mathcal{D}_\eta(a) = \eta_a$, then the coefficient $\prod_{i=1}^n \eta_{a_i}$ above is $\mathcal{D}_\eta^{\otimes n}(a_1, \ldots, a_n)$; i.e., the probability that a random length-n word drawn i.i.d. from $\mathcal{D}_\eta$ is equal to $(a_1, \ldots, a_n)$. From the definition of $\mathtt{P}(\pi)$ we further deduce

$$\mathtt{P}(\pi) \rho^{\otimes n} = \sum_{(a_1, \ldots, a_n)} \mathcal{D}_\eta^{\otimes n}(a_1, \ldots, a_n) \, |a_{\pi^{-1}(1)}, \ldots, a_{\pi^{-1}(n)}\rangle \langle a_1, \ldots, a_n|.$$

We immediately conclude that $\mathrm{tr}(\mathtt{P}(\boldsymbol{\pi}) \rho^{\otimes n})$ is equal to the sum over all π-invariant words $(a_1, \ldots, a_n)$ of $\mathcal{D}_\eta^{\otimes n}(a_1, \ldots, a_n)$. Recalling Definition 2.7, this is precisely given by the power sum polynomial $p_\pi(\eta_1, \ldots, \eta_d)$. Therefore

$$\mathrm{SW}^n_\rho(\lambda) = \dim(\lambda) \cdot \mathop{\mathbf{E}}_{\boldsymbol{\pi} \sim \mathfrak{S}_n} \left[\chi_\lambda(\boldsymbol{\pi}) \cdot p_\pi(\eta_1, \ldots, \eta_d) \right],$$

and the proposition now follows from Theorem 2.20. □

For the purposes of the testing lower bounds in this paper, the case of greatest interest to us is when $\rho = \frac{1}{d} I_{d \times d}$ is the *maximally mixed d-dimensional state*; i.e., the spectrum of ρ is the uniform distribution $\mathsf{Unif}_d = (\frac{1}{d}, \ldots, \frac{1}{d})$. This is also by far the most well-studied case in the literature:

Definition 2.25. The *Schur–Weyl* distribution with parameters n and d, which we denote SW_d^n, is the distribution on partitions $\lambda \vdash n$ of length at most d given by SW_ρ^n in the case that ρ is the maximally mixed state of dimension d. Equivalently, it is the distribution of $\mathrm{RSK}(\boldsymbol{a})$, where $\boldsymbol{a} \sim [d]^n$ is uniformly random.

Combining Proposition 2.23 and Proposition 2.11, together with the homogeneity of the Schur polynomials, we obtain the following known formula (cf. [CHW07, equation (26)]):

Proposition 2.26. $\mathrm{SW}_d^n(\lambda) = \dfrac{(\dim \lambda)^2}{n!} \cdot \dfrac{d^{\uparrow\lambda}}{d^n}$.

Notice that if n is held fixed and $d \to \infty$, the fraction $\frac{d^{\uparrow\lambda}}{d^n}$ tends to 1 and we obtain the Plancherel distribution (for $\mathfrak{S}_n$) on partitions described in Definition 2.17. This recovers the well-known fact that the Plancherel distribution is obtained by running the RSK algorithm on a uniformly random permutation (equivalently, a uniformly random word from $[0,1]^n$). We will write Planch_n for this distribution.

Remark 2.27. It is easy to see that $\mathrm{SW}_d^n(\lambda) = \frac{1}{d^n} \cdot \dim(\mathtt{p}_\lambda) \cdot \dim(\mathtt{q}_\lambda^d)$. From Remark 2.24, we see that there are $\dim(\mathtt{p}_\lambda) \cdot \dim(\mathtt{q}_\lambda^d)$ words $a \in [d]^n$ such that $\mathrm{RSK}(a) = \lambda$.

2.7 Asymptotic theory of the symmetric group

For small n, the exact distribution on partitions of n given by the Plancherel or Schur–Weyl distributions is not particularly easy to understand. As a result, a significant body of work has been devoted to showing asymptotic properties of these distributions as n grows large.

Let us focus first on the Plancherel measure. Perhaps the most basic thing one could ask for is the "typical" width and height of a diagram drawn from this distribution. Though either of these values could be as large as n, Hammersly [Ham72] showed that both values tend to concentrate around the same number $c \cdot \sqrt{n}$, for some constant c (later determined to be $c = 2$ [LS77, VK77]). Therefore, in order to put partitions of different values of n on equal footing, we can define scaled partitions as follows:

Definition 2.28. Let $\lambda \vdash n$ and recall Definition 2.3. Then $\overline{\lambda} : \mathbb{R} \to \mathbb{R}_+$ is defined as $\overline{\lambda}(x) := \lambda(\sqrt{n} \cdot x)/\sqrt{n}$, for all x.

Logan and Shepp [LS77] and Vershik and Kerov [VK77] independently proved the so-called "law of large numbers" for the Plancherel distribution, showing that when $\boldsymbol{\lambda} \sim \mathrm{Planch}_n$ and $n \to \infty$, the function $\overline{\boldsymbol{\lambda}}$ converges to $\Omega(x)$, the curve defined as

$$\Omega(x) := \begin{cases} \frac{2}{\pi}(x \arcsin \frac{x}{2} + \sqrt{4 - x^2}), & |x| \le 2, \\ |x| & |x| \ge 2. \end{cases}$$

This "ice cream cone"-shaped function is pictured in Figure 3 ($c = 0$ case). Though this curve is a limiting shape rather than the Russian notation of any Young diagram, it is useful to think of it as a continual analogue of a Young diagram, as per the following definition.

Definition 2.29. A *continual diagram* is a function $f : \mathbb{R} \to \mathbb{R}$ satisfying (i) f is 1-Lipschitz and (ii) $f(x) = |x|$ when $|x|$ is sufficiently large.

This definition originates in the paper of [Ker93a].

More recently, Kerov [Ker93b] showed a "central limit theorem" for the Plancherel measure, characterizing the deviation of a random Young diagram from the curve $\Omega(x)$ by a certain Gaussian

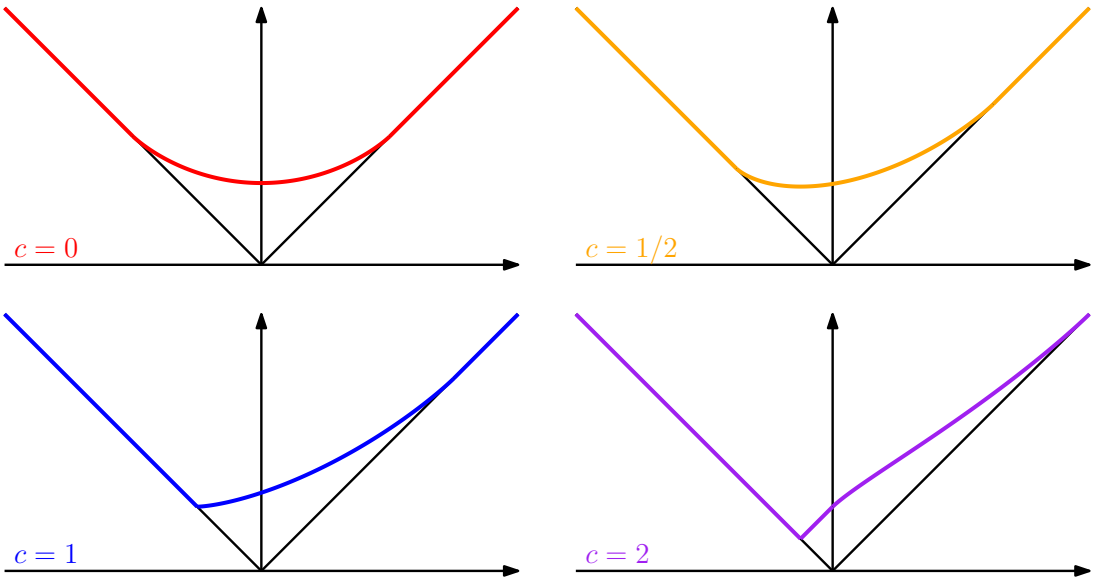


Figure 3: The Biane limiting curves Ω_c. The $c = 0$ case corresponds to the function $\Omega(x)$.

process. A second proof of this result, also by Kerov, was given in the paper of Ivanov and Olshanski [IO02]. Much of our work is based on the techniques of this paper.

Subsequent studies revealed that a similar state of affairs exists for the Schur–Weyl SW_d^n distribution, though in this case the features of a "typical" $\boldsymbol{\lambda} \sim \mathrm{SW}_d^n$ depend on the ratio $c := \frac{\sqrt{n}}{d}$. Biane [Bia01] extended the Plancherel law of large numbers to the Schur–Weyl distribution in the case when c is a fixed constant and $n, d \to \infty$. In this case, for a random $\boldsymbol{\lambda} \sim \mathrm{SW}_d^n$, the function $\overline{\boldsymbol{\lambda}}$ will approach a certain limiting curve Ω_c, specified as follows:

Theorem 2.30 ([Bia01])**.** *Fix an absolute constant $c > 0$ and assume $n, d \to \infty$ with $\frac{\sqrt{n}}{d} \to c$. Then*

$$\mathop{\mathbf{Pr}}_{\boldsymbol{\lambda} \sim \mathrm{SW}_d^n} \left[\|\overline{\boldsymbol{\lambda}} - \Omega_c\|_\infty \geq \epsilon \right] \to 0,$$

where Ω_c is the continual diagram defined as follows:

$$\Omega_0(x) = \Omega(x);$$

$$\Omega_{c \in (0,1)}(x) = \begin{cases} \frac{2}{\pi}\left(x \arcsin(\frac{x+c}{2\sqrt{1+cx}}) + \frac{1}{c}\arccos(\frac{2+cx-c^2}{2\sqrt{1+cx}}) + \frac{\sqrt{4-(x-c)^2}}{2} \right) & \text{if } |x-c| \leq 2, \\ |x| & \text{otherwise;} \end{cases}$$

$$\Omega_{c=1}(x) = \begin{cases} \frac{x+1}{2} + \frac{1}{\pi}\left((x-1)\arcsin(\frac{x-1}{2}) + \sqrt{4-(x-1)^2} \right) & \text{if } |x-1| \leq 2, \\ |x| & \text{otherwise;} \end{cases}$$

$$\Omega_{c>1}(x) = \begin{cases} x + \frac{2}{c} & \text{if } x \in (\frac{-1}{c}, c-2) \\ \frac{2}{\pi}\left(x \arcsin(\frac{x+c}{2\sqrt{1+cx}}) + \frac{1}{c}\arccos(\frac{2+cx-c^2}{2\sqrt{1+cx}}) + \frac{\sqrt{4-(x-c)^2}}{2} \right) & \text{if } |x-c| \leq 2, \\ |x| & \text{otherwise.} \end{cases}$$

These curves are pictured for various values of c in Figure 3 (which we have reproduced from [Mél10a]). Meliot [Mél10a, Mél10b] has extended Kerov's central limit theorem to the Schur–Weyl distribution, characterizing the fluctuations of $\overline{\boldsymbol{\lambda}}$ around the limiting curves given by Biane.

One consequence is these results is that when $n = o(d^2)$, the function $\overline{\boldsymbol{\lambda}}$ converges to the ice cream cone curve $\Omega(x)$ from above. This fact is a manifestation of the discussion at the end of Section 2.6 concerning SW_d^n tending to Planch_n as $d \to \infty$. Indeed, Childs et al. [CHW07] showed that when $n = o(d)$, the two distributions are statistically indistinguishable (from which the lower bound in Theorem 1.9 follows via the triangle inequality $d_{\mathrm{TV}}(\mathrm{SW}_r^n, \mathrm{SW}_{2r}^n) \leq d_{\mathrm{TV}}(\mathrm{SW}_r^n, \mathrm{Planch}_n) + d_{\mathrm{TV}}(\mathrm{Planch}_n, \mathrm{SW}_{2r}^n)$).

We close this section by recording some simple concentration bounds on the width and length of $\boldsymbol{\lambda} \sim \mathrm{SW}_d^n$. They are not as precise as what is suggested by the above limit theorems, but they have the advantage of giving concrete error bounds. We follow a simple line of argument similar to that in [Rom14, Lemma 1.5].

Proposition 2.31. *Let $\boldsymbol{\lambda} \sim \mathrm{SW}_d^n$. For every $B \in \mathbb{Z}^+$ we have $\mathbf{Pr}[\boldsymbol{\lambda}_1 \geq B] \leq \left(\frac{(1+B/d)e^2 n}{B^2}\right)^B$. The same bound holds for $\mathbf{Pr}[\boldsymbol{\lambda}'_1 \geq B]$.*

Proof. By Theorem 2.14, $\mathbf{Pr}[\boldsymbol{\lambda}_1 \geq B]$ (respectively, $\mathbf{Pr}[\boldsymbol{\lambda}'_1 \geq B]$) is equal to the probability that a uniformly random word from $[d]^n$ contains a weakly increasing (respectively, strongly increasing) subsequence of length exactly B. As weakly increasing subsequences are more probable than strongly increasing ones, it suffices to bound

$$\mathbf{Pr}[\boldsymbol{\lambda}_1 \geq B] \leq \left(\frac{(1+B/d)e^2 n}{B^2}\right)^B.$$

Letting $\boldsymbol{S}$ denote the number of weakly increasing subsequences of length B in a random word we have

$$\mathbf{Pr}[\boldsymbol{\lambda}_1 \geq B] \leq \mathbf{E}[\boldsymbol{S}] = \binom{n}{B} \cdot \frac{c}{d^B},$$

where c is the number of words in $[d]^B$ which are weakly increasing. Evidently c also equals the number of "weak d-compositions of B", which [Sta11, Chapter 1.2] is $\binom{d-1+B}{B} \leq \binom{d+B}{B}$. We conclude

$$\mathbf{Pr}[\boldsymbol{\lambda}_1 \geq B] \leq \binom{n}{B} \cdot \frac{\binom{d+B}{B}}{d^B} \leq \frac{\left(\frac{en}{B}\right)^B \left(\frac{(1+B/d)ed}{B}\right)^B}{d^B} = \left(\frac{(1+B/d)e^2 n}{B^2}\right)^B,$$

as needed. □

2.8 Polynomial algebras

We have already discussed the power sum and Schur polynomials, which are elements of the $\mathbb{C}$-algebra Λ of symmetric polynomials in indeterminates $x_1, x_2, \ldots$.[8] Important to our work will be a closely related polynomial algebra Λ^*, the algebra of *shifted symmetric* polynomials, formally introduced introduced in [OO98b]. This algebra consists of those polynomials which are symmetric in the "shifted" indeterminates $\widetilde{x}_i := x_i - i + c$, where c is any fixed constant. (The definition does not depend on the constant c.) When we view the inputs to the shifted symmetric functions $x_1, x_2, \ldots$ as the values $\lambda_1, \lambda_2, \ldots$ of a partition λ, the result is (isomorphic to) *Kerov's algebra* of polynomial functions on the set of Young diagrams, also known as the *algebra of observables*

[8] Strictly speaking, these are *families* of bounded-degree polynomials, one for each number of indeterminates, which are *stable* in the sense that $p_\lambda(x_1, \ldots, x_d, 0) = p_\lambda(x_1, \ldots, x_d)$, and similarly for s_λ. See, e.g., [Mac95] for a formal definition via projective limits.

of diagrams. In a nutshell, the importance of this algebra is that, on one hand, it still contains polynomials that are similar to "power sums" or "moments" of the λ_i's; and, on the other hand, it is easier to compute their expected value under SW^n_ρ distributions.

We will need to study several families of observables/shifted symmetric polynomials, and their relationships:

Definition 2.32. The following polynomials are known to be elements of Λ^*. (We describe the first four as observables of Young diagrams.)

- For $k \geq 1$,

$$p^*_k(\lambda) \coloneqq \sum_{i=1}^{d(\lambda)} \Big((a^*_i)^k - (-b^*_i)^k \Big) = \sum_{i=1}^{\infty} \Big((\lambda_i - i + \tfrac{1}{2})^k - (-i+\tfrac{1}{2})^k \Big).$$

  These are the most basic polynomials on Young diagrams, giving the "moments" of the coordinates. For more information on them see [IO02], where they are introduced (in equation (1.4)) under the notation $p_k(\lambda)$. We use the notation $p^*_k(\lambda)$ to distinguish them from the ordinary power sum symmetric polynomials. It is obvious from the second definition above that the p^*_k polynomials are in Λ^*. In fact they are algebraically independent, and they generate Λ^*.

- For $k \geq 0$, the kth *content sum* polynomial is $c_k(\lambda) \coloneqq \sum_{\square \in [\lambda]} c(\square)^k$. Although these polynomials are quite natural, we will have little occasion to use them. The fact that they are in Λ^* was proven in [KO94].

- For $k \geq 2$,

$$\widetilde{p}_k(\lambda) \coloneqq k(k-1) \int_{-\infty}^{\infty} x^{k-2} \sigma(x)\, dx,$$

  where $\sigma(x) \coloneqq \frac{1}{2}(\lambda(x) - |x|)$. These polynomials were introduced and shown to be algebraically independent generators of Λ^* in [IO02, Section 2]. They can shown to be the "moments of the local extrema of $\lambda(x)$", and are also useful for studying continual diagrams. We use them only briefly, to pass between the p^*_k polynomials and $p^\sharp_k$ polynomials defined below.

- For $\lambda \vdash n$ and $\mu \vdash k$, the *central characters* are defined by

$$p^\sharp_\mu(\lambda) = \begin{cases} n^{\downarrow k} \cdot \frac{\chi_\lambda(\mu \cup 1^{n-k})}{\dim(\lambda)} & \text{if } n \geq k, \\ 0 & \text{if } n < k. \end{cases}$$

  where $\mu \cup 1^{n-k}$ denotes the partition $(\mu, 1, 1, \ldots, 1) \vdash n$. In case $\mu = (k)$ we simply write $p^\sharp_k(\lambda)$. Note that we are somewhat unexpectedly applying the character χ_λ to (an extension of) μ, and not the other way around. The advantage of the $p^\sharp_\mu$ polynomials is that, by virtue of them being characters of the symmetric group (up to some normalizations), their expectations under SW^n_ρ can be easily calculated exactly, as we will see below. A disadvantage is that, by virtue of them being characters of the symmetric group, explicit formulas for them are famously quite complex [Las08, Fér10] (though in Section 2.8.1 we will mention a formula that allows one to compute $p^\sharp_k$ for small k fairly easily). Wassermann [Was81, III.6] showed that the $p^\sharp_k$ polynomials are in Λ^*, and in fact [VK81, KO94, OO98b] more generally the polynomials $p^\sharp_\mu$ form a *linear* basis of Λ^*.

- For $\mu \vdash k$, the *shifted Schur* polynomial in indeterminates $x_1, \ldots, x_d$ is

$$s^*_\mu(x_1, \ldots, x_d) = \frac{\det\Big((x_i - i + d)^{\downarrow(d+\lambda_j - j)}\Big)_{ij}}{\det\Big((x_i - i + d)^{\downarrow(d-j)}\Big)_{ij}} \quad \text{if } \ell(\mu) \le d, \text{ else } 0.$$

  These polynomials are the shifted analogues of the Schur polynomials (cf. Theorem 2.13). They were introduced by Okounkov and Olshanski [OO98b], and are similar to the earlier-defined "factorial Schur functions" (see, e.g., [Mac95, I.3.20–21]), but with the advantage that they are *stable*—i.e., $s^*_\mu(x_1, \ldots, x_d, 0) = s^*_\mu(x_1, \ldots, x_d)$. They arise for us because they can sometimes be used to express the ratio of two Schur functions (see the "Binomial Formula" Theorem 4.6). To analyze them, we will use the following "shifted analogue" of Theorem 2.20, proved in [OO98b, Theorem 8.1], [IK01, Theorem 9.1] (see also [Mél10b, p.25]):

  Theorem 2.33. *For $\mu \vdash k$, let us think of the central character polynomial $p^\sharp_\mu$ not as an observable of Young diagrams (applied to $\lambda_1, \ldots, \lambda_d$) but as a shifted symmetric polynomial in indeterminates $x_1, \ldots, x_d$. In the context of Fourier analysis over the group $G = \mathfrak{S}_k$, for each fixed $x \in \mathbb{C}^d$ we may think of $p^\sharp_{(\cdot)}(x) \coloneqq \pi \mapsto p^\sharp_\pi(x)$ as a class function. Then its Fourier coefficients are given by*

$$\widehat{p^\sharp_{(\cdot)}(x)}(\mu) = s^*_\mu(x).$$

  (Note that give the determinantal definition of the shifted Schur polynomials, one may alternatively take this Theorem as a definition of the shifted symmetric polynomials $p^\sharp_\mu(x)$.)

As mentioned, the $p^\sharp_\mu$ polynomials are especially important for us as because there is a simple expression for their expectation under any Schur–Weyl distribution. This is the subject of our next proposition.

Proposition 2.34. *Let ρ be a $d \times d$ density matrix with eigenvalues $\eta_1, \ldots, \eta_d$, and let $\mu \vdash k$. Then*

$$\mathop{\mathbf{E}}_{\boldsymbol{\lambda} \sim \mathrm{SW}^n_\rho}[p^\sharp_\mu(\boldsymbol{\lambda})] = n^{\downarrow k} \cdot p_\mu(\eta_1, \ldots, \eta_d).$$

Proof. It's immediate from the definitions that both sides are 0 if $n < k$, so we assume $n \ge k$. Applying Proposition 2.23 and the definition of $p^\sharp_\mu$ we obtain

$$\begin{aligned}\mathop{\mathbf{E}}_{\boldsymbol{\lambda} \sim \mathrm{SW}^n_\rho}[p^\sharp_\mu(\boldsymbol{\lambda})] &= n^{\downarrow k} \cdot \sum_{\lambda \vdash n} s_\lambda(\eta_1, \ldots, \eta_d) \cdot \chi_\lambda(\mu \cup 1^{n-k}) \\ &= n^{\downarrow k} \cdot p_{\mu \cup 1^{n-l}}(\eta_1, \ldots, \eta_d),\end{aligned}$$

where the second equation is from Theorem 2.20. But $p_{\mu \cup 1^{n-l}}(\eta_1, \ldots, \eta_d) = p_\mu(\eta_1, \ldots, \eta_d)$, since the two quantities differ only by factors of $p_1(\eta_1, \ldots, \eta_d) = \eta_1 + \cdots + \eta_d = 1$. □

Note that in the case of $\eta_1 = \ldots = \eta_d = 1/d$, we have that $p_\mu(\eta_1, \ldots, \eta_d) = d^{\ell(\mu)-k}$. This gives us the following important corollary:

Corollary 2.35. *Let $\mu \vdash k$. Then* $\mathop{\mathbf{E}}_{\boldsymbol{\lambda} \sim \mathrm{SW}^n_d}[p^\sharp_\mu(\boldsymbol{\lambda})] = n^{\downarrow k} \cdot d^{\ell(\mu)-k}$.

2.8.1 Working with the $p^\sharp_\mu$ polynomials

As we will be working heavily with the $p^\sharp_\mu$ polynomials, let us describe them further. We begin with the simpler case of the $p^\sharp_k$ polynomials. Let us see how these polynomials can be written in terms of the p^*_k polynomials. From [Was81, III.6] (cf. [IO02, Proposition 3.3]) we have the following identity using generating functions:

$$p^\sharp_k = [t^{k+1}]\left\{-\frac{1}{k}\prod_{j=1}^{k}(1-(j-\tfrac{1}{2})t)\cdot\exp\left(\sum_{j=1}^{\infty}\frac{p^*_j t^j}{j}(1-(1-kt)^{-j})\right)\right\}.$$

One may rewrite this (cf. [IO02, (3.3)]) as

$$p^\sharp_k = [t^{k+1}]\left\{-\frac{1}{k}\prod_{j=1}^{k}(1-(j-\tfrac{1}{2})t)\cdot\sum_{i=0}^{\infty}\frac{(-1)^i}{i!}Q_k(t)^i\right\}, \tag{7}$$

where

$$Q_k(t)=\sum_{m=1}^{\infty}Q_{k,m}t^{m+1},\quad Q_{k,m}=\tfrac{1}{1}\tbinom{m}{0}k^m p^*_1+\tfrac{1}{2}\tbinom{m}{1}k^{m-1}p^*_2+\tfrac{1}{3}\tbinom{m}{2}k^{m-2}p^*_3+\cdots+\tfrac{1}{m}\tbinom{m}{m-1}kp^*_m. \tag{8}$$

It follows that in (7) we may restrict the sum on i to the range between 0 and $\frac{k+1}{2}$, and in (8) we can restrict the sum on m to the range between 1 and k. We thereby obtain a relatively simple finitary method for expressing $p^\sharp_k$'s polynomials in terms of p^*_j's. In particular, we can deduce

$$p^\sharp_1=p^*_1,\qquad p^\sharp_2=p^*_2,\qquad p^\sharp_3=p^*_3-\tfrac{3}{2}(p^*_1)^2+\tfrac{5}{4}p^*_1,\qquad p^\sharp_4=p^*_4-4p^*_2p^*_1+\tfrac{11}{2}p^*_2. \tag{9}$$

As observed in [IO02, Proposition 3.4], we can also deduce that in general,

$$p^\sharp_k = p^*_k+\Big\{\text{polynomial in } p^*_1,\dots,p^*_{k-1} \text{ of gradation at most } k-1\Big\}, \tag{10}$$

where *gradation* refers to the canonical grading in which $\prod_i p^*_{\lambda_i}$ has gradation $|\lambda|$. We can of course inductively invert this relationship, deducing that

$$p^*_k = p^\sharp_k+\Big\{\text{polynomial in } p^\sharp_1,\dots,p^\sharp_{k-1} \text{ of gradation at most } k-1\Big\}. \tag{11}$$

For example,

$$p^*_1=p^\sharp_1,\qquad p^*_2=p^\sharp_2,\qquad p^*_3=p^\sharp_3+\tfrac{3}{2}(p^\sharp_1)^2-\tfrac{5}{4}p^\sharp_1,\qquad p^*_4=p^\sharp_4+4p^\sharp_2p^\sharp_1-\tfrac{11}{2}p^\sharp_2. \tag{12}$$

Recall that the more general $p^\sharp_\tau$ polynomials (for $\tau\in\mathrm{Par}$) are known to linearly generate the algebra of observables. This means that any product $p^\sharp_{\mu_1}p^\sharp_{\mu_2}$ can be converted to a linear combination of $p^\sharp_\tau$'s. In particular, if we applied this conversion in (12) we would get linear expressions for the "low-degree moments of Young diagrams" (i.e., the p^*_j's) in terms of $p^\sharp_\tau$'s; we could then compute the expectation of these, under any Schur–Weyl distribution, using Proposition 2.34.

We are therefore interested in the *structure constants* $f^\tau_{\mu_1\mu_2}$ of Λ^* in the basis $\{p^\sharp_\tau\}$; i.e., the numbers such that

$$p^\sharp_{\mu_1}p^\sharp_{\mu_2}=\sum_{\tau\in\mathrm{Par}}f^\tau_{\mu_1\mu_2}p^\sharp_\tau.$$

These were first determined by Ivanov and Kerov [IK01] in terms of the algebra of *partial permutations*. We quote the following formulation from [IO02, Proposition 4.5]:

Proposition 2.36. *Let $\tau, \mu_1, \mu_2 \in \mathrm{Par}$. Fix a set R of cardinality $|\tau|$ and a permutation $w : R \to R$ of cycle type τ. Then*

$$f^{\tau}_{\mu_1\mu_2} = \frac{z_{\mu_1} z_{\mu_2}}{z_\tau} g^{\tau}_{\mu_1\mu_2},$$

where $g^{\tau}_{\mu_1\mu_2}$ equals the number of quadruples (R_1, w_1, R_2, w_2) such that:

1. *$R_1 \subseteq R, \quad R_2 \subseteq R, \quad R_1 \cup R_2 = R$;*
2. *$|R_i| = |\mu_i|$ and $w_i : R_i \to R_i$ is a permutation of cycle type μ_i, for $i = 1, 2$;*
3. *$\overline{w}_1 \overline{w}_2 = w$, where $\overline{w}_i : R \to R$ denotes the natural extension of w_i from R_i to the whole of R.*

We present an equivalent formulation we have found to be more convenient. We omit its straightforward combinatorial deduction from Proposition 2.36.

Corollary 2.37. *Let*

$$C^{t}_{r_1 r_2} := \frac{r_1! r_2!}{(t - r_1)!(t - r_2)!(r_1 + r_2 - t)!}$$

if the positive integers r_1, r_2, t satisfy $r_1, r_2 \leq t \leq r_1 + r_2$, and let $C^{t}_{r_1 r_2} := 0$ otherwise. Then for $\mu \vdash r_1$, $\nu \vdash r_2$, $\tau \vdash t$,

$$f^{\tau}_{\mu\nu} = C^{t}_{r_1 r_2} \cdot \mathop{\mathbf{Pr}}_{\boldsymbol{w}_1, \boldsymbol{w}_2} \left[\overline{\boldsymbol{w}}_1 \overline{\boldsymbol{w}}_2 \text{ has cycle type } \tau \right],$$

where $\boldsymbol{w}_1$ is a uniformly random permutation on $\{1, \ldots, r_1\}$ of cycle type μ, and $\boldsymbol{w}_2$ is a uniformly random permutation on $\{t - r_2 + 1, \ldots, t\}$ of cycle type ν.

As very simple examples, we can compute

$$(p^\sharp_1)^2 = p^\sharp_{(1,1)} + p^\sharp_1, \qquad p^\sharp_2 p^\sharp_1 = p^\sharp_{(2,1)} + 2p^\sharp_2, \qquad (p^\sharp_2)^2 = p^\sharp_{(2,2)} + 4p^\sharp_3 + 2p^\sharp_{(1,1)}. \tag{13}$$

Substituting these into (9), we obtain the formulas

$$p^*_1 = p^\sharp_1, \qquad p^*_2 = p^\sharp_2, \qquad p^*_3 = p^\sharp_3 + \tfrac{3}{2} p^\sharp_{(1,1)} + \tfrac{1}{4} p^\sharp_1, \qquad p^*_4 = p^\sharp_4 + 4p^\sharp_{(2,1)} + \tfrac{5}{2} p^\sharp_2, \tag{14}$$

which will be useful to us later.

Given the formula for the structure constants, it's not hard to show that

$$p^\sharp_\mu p^\sharp_\nu = p^\sharp_{\mu \cup \nu} + \Big\{ \text{linear combination of } p^\sharp_\tau\text{'s with } |\tau| < |\mu \cup \nu| \Big\},$$

where $\mu \cup \nu$ denotes the partition formed by joining the parts of μ and ν and sorting them in nonincreasing order (i.e., $m_w(\mu \cup \nu) = m_w(\mu) + m_w(\nu)$). In fact, we will require a stronger statement, based on the following notion introduced in [IK01]:

Definition 2.38. For a partition $\lambda \in \mathrm{Par}$, its *weight* is defined to be $\mathrm{wt}(\lambda) = |\lambda| + \ell(\lambda)$.

Now Śniady [Śni06, Corollary 3.8] proved:

Proposition 2.39. $p^\sharp_\mu p^\sharp_\nu = p^\sharp_{\mu \cup \nu} + \Big\{ \text{linear combination of } p^\sharp_\tau\text{'s with } \mathrm{wt}(\tau) \leq \mathrm{wt}(\mu) + \mathrm{wt}(\nu) - 2 \Big\}$.

3 The empirical Young diagram algorithm

The empirical Young diagram (EYD) algorithm works as follows:

The EYD algorithm. Given $\rho^{\otimes n}$:

1. Sample $\boldsymbol{\lambda} \sim \mathrm{SW}_\rho^n$.
2. Output $\underline{\boldsymbol{\lambda}} := (\boldsymbol{\lambda}_1/n, \ldots, \boldsymbol{\lambda}_d/n)$.

This algorithm has, either implicitly or explicitly, arisen in several independent research threads. The first was the work of Alicki, Rudnicki, and Sadowski [ARS88], who showed that if ρ has eigenvalues $\eta_1 \geq \ldots \geq \eta_d$, then $\underline{\boldsymbol{\lambda}} \to \eta$ as $n \to \infty$, and furthermore sketched a central limit theorem for the fluctuations. Ten years later, Keyl and Werner [KW01] independently reproved the first part of this result (and showed an "error rate" for the EYD algorithm which, for any fixed d, decreases exponentially in n); they also explicitly suggested the EYD algorithm for spectrum estimation. Further independent work, developing the research on the "Gaussian Unitary Ensemble" nature of the fluctuations, was performed by Its–Tracy–Widom, Houdré and coauthors, and others [ITW01, Lit08, HX13]

3.1 The upper bound

Following Keyl and Werner's paper [KW01], a short, simplified proof of correctness containing explicit error bounds was discovered in [HM02]. A small bug in their derivation was corrected by [CM06], whose Corollary 1 states:

Theorem 3.1. *Let ρ be a mixed state with eigenvalues $\eta_1 \geq \ldots \geq \eta_d$. Let S be any set of partitions of n, and set $d_{\mathrm{KL}} := \min_{\lambda \in S} d_{\mathrm{KL}}(\underline{\lambda}, \eta)$. Then*

$$\Pr_{\boldsymbol{\lambda} \sim \mathrm{SW}_\rho^n}[\boldsymbol{\lambda} \in S] \leq (n+1)^{d(d+1)/2} \cdot e^{-n \cdot d_{\mathrm{KL}}}.$$

If we apply Theorem 3.1 with the set of partitions $S = \{\lambda \vdash n \mid d_{\mathrm{TV}}(\underline{\lambda}, \eta) > \epsilon\}$ and use Pinsker's inequality, we get the following corollary:

Corollary 3.2. *Let ρ be a mixed state with eigenvalues $\eta_1 \geq \ldots \geq \eta_d$. Then*

$$\Pr_{\boldsymbol{\lambda} \sim \mathrm{SW}_\rho^n}[d_{\mathrm{TV}}(\underline{\boldsymbol{\lambda}}, \eta) > \epsilon] \leq (n+1)^{d(d+1)/2} \cdot e^{-2n\epsilon^2}.$$

In particular, $O(d^2/\epsilon^2) \cdot \log(d/\epsilon) \cdot \log(1/\delta)$ samples are sufficient to output an estimate $\underline{\lambda}$ satisfying $d_{\mathrm{TV}}(\underline{\lambda}, \eta) \leq \epsilon$ with probability at least $1 - \delta$.

This means that any unitarily invariant property of mixed states is testable with $O(d^2/\epsilon^2) \cdot \log(d/\epsilon)$ copies.

We now give a simplified proof of Theorem 3.1. This will largely follow the outline of the proof found in [HM02, CM06], except we will reinterpret their majorizing step in light of the RSK algorithm.

Proof of Theorem 3.1. Define the probability distribution $\mathcal{D} = (\eta_1, \ldots, \eta_d)$. For a fixed partition $\lambda \in S$, Remark 2.24 shows that upper-bounding $\mathrm{SW}_\rho^n(\lambda)$ is equivalent to upper-bounding $\mathbf{Pr}_{\boldsymbol{a} \sim \mathcal{D}^{\otimes n}}[\mathrm{RSK}(\boldsymbol{a}) = \lambda]$. By Proposition 2.15, $\mathrm{RSK}(a) = \lambda$ only if λ *majorizes* $c(a)$.

By Remark 2.27, there are exactly $\dim(\mathtt{p}_\lambda) \cdot \dim(\mathtt{q}^d_\lambda)$ words $a \in [d]^n$ for which $\mathrm{RSK}(a) = \lambda$. By the majorizing step, the probability that such an a is drawn from $\mathcal{D}^{\otimes n}$ is

$$\prod_i \eta_i^{c_i(a)} \leq \prod_i \eta_i^{\lambda_i}.$$

From this point on, the rest of the argument is as in [HM02, CM06]. Recall the well-known upper bounds (cf. [Chr06, Equations (1.21) and (1.22)])

$$\dim(\mathtt{p}_\lambda) \leq \frac{n!}{\prod_i \lambda_i!}, \qquad \dim(\mathtt{q}^d_\lambda) \leq (n+1)^{d(d-1)/2}.$$

Thus, we can upper-bound $\mathbf{Pr}_{\boldsymbol{a} \sim \mathcal{D}^{\otimes n}}[\mathrm{RSK}(\boldsymbol{a}) = \lambda]$ by

$$(n+1)^{d(d-1)/2} \cdot \frac{n!}{\prod_i \lambda_i!} \cdot \prod_i \eta_i^{\lambda_i} \leq (n+1)^{d(d-1)/2} \cdot \exp(-n \cdot d_{\mathrm{KL}}(\underline{\lambda}, \eta)).$$

To recover Theorem 3.1, we now union bound over all $\lambda \in S$, of which there are at most $(n+1)^d$. □

3.2 The lower bound

Our main result of this section is that Corollary 3.2 is nearly tight, even when ρ is the maximally mixed state. In particular, we show the following lower bound:

Theorem 3.3. *There is a $\delta > 0$ such that for sufficiently small values of ϵ,*

$$\mathop{\mathbf{Pr}}_{\boldsymbol{\lambda} \sim \mathrm{SW}^n_d}[d_{\mathrm{TV}}(\underline{\boldsymbol{\lambda}}, \mathsf{Unif}_d) > \epsilon] \geq \delta$$

unless $n = \Omega(d^2/\epsilon^2)$.

We will split the lower bound into two cases.

Theorem 3.4. *For every constant $C > 0$, there are constants $\delta, \epsilon > 0$ such that*

$$\mathop{\mathbf{Pr}}_{\boldsymbol{\lambda} \sim \mathrm{SW}^n_d}[d_{\mathrm{TV}}(\underline{\boldsymbol{\lambda}}, \mathsf{Unif}_d) > \epsilon] \geq \delta$$

when $n < Cd^2$ and d is sufficiently large.

Theorem 3.5. *There are absolute constants $C > 0$ and $0 < \delta < 1$ such that*

$$\mathop{\mathbf{Pr}}_{\boldsymbol{\lambda} \sim \mathrm{SW}^n_d}[d_{\mathrm{TV}}(\underline{\boldsymbol{\lambda}}, \mathsf{Unif}_d) > \epsilon] \geq \delta$$

when $n \geq Cd^2$, unless $n = \Omega(d^2/\epsilon^2)$.

To prove Theorem 3.3, let C and δ_1 be the constants in Theorem 3.5. Apply Theorem 3.4 with the value of C, and let δ_2 and ϵ_0 be the resulting constants. Set $\delta := \min\{\delta_1, \delta_2\}$. Then we see that for all $\epsilon \leq \epsilon_0$,

$$\mathop{\mathbf{Pr}}_{\boldsymbol{\lambda} \sim \mathrm{SW}^n_d}[d_{\mathrm{TV}}(\underline{\boldsymbol{\lambda}}, \mathsf{Unif}_d)) > \epsilon] \geq \delta$$

unless $n = \Omega(d^2/\epsilon^2)$, giving Theorem 3.3.

Theorem 3.4 might appear somewhat superfluous, as Theorem 3.5 already proves the lower bound for sufficiently large values of n (i.e., $n \geq Cd^2$), and intuitively having fewer copies of ρ

shouldn't improve the performance of the EYD algorithm. However, this intuition, though it may be true in some approximate sense, is false in general: there are regimes of state estimation where the performance of the EYD algorithm does *not* increase monotonically with the value of n. For example, if n is a multiple of d, then when $\boldsymbol{\lambda} \sim \mathrm{SW}_d^n$, $\underline{\boldsymbol{\lambda}}$ will equal Unif_d with some nonzero probability. On the other hand, a random $\boldsymbol{\lambda} \sim \mathrm{SW}_d^{n+1}$ will never be uniform, because $n+1$ is not a multiple of d. Thus, decreasing the value of n can sometimes help (according to some performance metrics), and this shows why we need Theorem 3.4 to supplement Theorem 3.5.

The proof of Theorem 3.4 is quite technical, and we defer it to Section 7. Our proof of Theorem 3.5 is simpler and appears below. It is a good illustration of the basic technique of using polynomial functions on Young diagrams. The intuition behind the proof is as follows: By the (traceless) Gaussian Unitary Ensemble fluctuations predicted in [ITW01], we expect that for $\boldsymbol{\lambda} \sim \mathrm{SW}_d^n$, the empirical distribution $\underline{\boldsymbol{\lambda}}$ will deviate from Unif_d by roughly $\Theta(1/\sqrt{n})$ in each coordinate. This will yield total variation distance $\Theta(d/\sqrt{n})$, necessitating $n \geq \Omega(d^2/\epsilon^2)$ to achieve $d_{\mathrm{TV}}(\underline{\boldsymbol{\lambda}}, \mathsf{Unif}_d) \leq \epsilon$. Actually analyzing the precise rate of convergence to Gaussian fluctuations in terms of n is difficult, and is overkill anyway; instead, we use the Fourth Moment Method to lower bound the fluctuations.

Proof of Theorem 3.5. Our goal is to show that for $n \geq 10^{10} d^2$, with 1% probability over a random $\boldsymbol{\lambda} \sim \mathrm{SW}_d^n$, at least $\frac{d}{200}$ coordinates $i \in [d]$ satisfy

$$\left| \boldsymbol{\lambda}_i - \frac{n}{d} \right| \geq \frac{\sqrt{n}}{1000}.$$

When this event occurs,

$$\begin{aligned} d_{\mathrm{TV}}(\underline{\boldsymbol{\lambda}}, \mathsf{Unif}_d) = \frac{1}{2} \cdot \sum_{i=1}^{d} \left| \frac{\boldsymbol{\lambda}_i}{n} - \frac{1}{d} \right| &= \frac{1}{2} \cdot \sum_{i=1}^{d} \frac{1}{n} \cdot \left| \boldsymbol{\lambda}_i - \frac{n}{d} \right| \\ &\geq \frac{1}{2} \cdot \frac{d}{200} \cdot \frac{1}{n} \cdot \frac{\sqrt{n}}{1000} = \frac{1}{400000} \cdot \frac{d}{\sqrt{n}}, \end{aligned}$$

which is bigger than ϵ unless $n = \Omega(d^2/\epsilon^2)$. Showing this will prove Theorem 3.5 with the parameters $C = 10^{10}$ and $\delta = .01$.

To begin, let us define a family of polynomials.

Definition 3.6. Given $k \geq 1$ and $c \in \mathbb{R}$, we define $p^*_{k,c}(\lambda) := \sum_{i=1}^{\infty} (\lambda_i - i - c)^k - (-i - c)^k$.

This generalizes the definition of the p^*_k polynomials, as $p^*_{k,-\frac{1}{2}} = p^*_k$.

Fact 3.7. *Let $c \in \mathbb{R}$. Then*

- $p^*_{2,c} = (-2c - 1)p^\sharp_1 + p^\sharp_2$, *and*
- $p^*_{4,c} = (-4c^3 - 6c^2 - 4c - 1)p^\sharp_1 + (6c^2 + 6c + 4)p^\sharp_2 + (-6c - 3)p^\sharp_{(1,1)} + (-4c - 2)p^\sharp_3 + 4p^\sharp_{(2,1)} + p^\sharp_4$.

Proof. By explicit computation, one can check that

$$p^*_{2,c} = 2(-c - \tfrac{1}{2})p^*_1 + p^*_2, \qquad p^*_{4,c} = 4(-c - \tfrac{1}{2})^3 p^*_1 + 6(-c - \tfrac{1}{2})^2 p^*_2 + 4(-c - \tfrac{1}{2})p^*_3 + p^*_4.$$

(Indeed, it's not hard to show that in general, $p^*_{k,c} = \sum_{j=1}^{k} \binom{k}{j} (-c - \frac{1}{2})^{k-j} p^*_j$.) The claim now follows from (14). ∎

For any c, these formulas allow us to compute the expected value of $p^*_{2,c}$ and $p^*_{4,c}$ over a random $\boldsymbol{\lambda} \sim \mathrm{SW}^n_d$, by using Corollary 2.35. Furthermore, for any k and d, $\sum_{i=1}^d (-i-c)^k$ is a constant which doesn't depend on $\boldsymbol{\lambda}$. Combining these two facts allows us to compute average value over a random $\boldsymbol{\lambda} \sim \mathrm{SW}^n_d$ of $\sum_{i=1}^d (\boldsymbol{\lambda}_i - i - c)^k$, for $k = 2, 4$. In particular, we are interested in computing this expectation when $c = \frac{n}{d}$. Write $\boldsymbol{L}_i \coloneqq \boldsymbol{\lambda}_i - i - \frac{n}{d}$. Then

$$\mathop{\mathbf{E}}_{\boldsymbol{\lambda} \sim \mathrm{SW}^n_d} \left[\sum_{i=1}^d \boldsymbol{L}_i^2 \right] = -\tfrac{n}{d} + nd + \tfrac{d^3}{3} + \tfrac{d^2}{2} + \tfrac{d}{6} \geq -\tfrac{n}{d} + nd \geq \tfrac{3nd}{4}, \tag{15}$$

where in the last step we used the fact that $n/d \leq nd/4$ because $d \geq 2$.

Similarly, as $n \geq 10^{10} d^2 \geq d^2$, we can use the bound

$$\begin{aligned}
\mathop{\mathbf{E}}_{\boldsymbol{\lambda} \sim \mathrm{SW}^n_d} \left[\sum_{i=1}^d \boldsymbol{L}_i^4 \right] &= 2n - \tfrac{d}{30} - \tfrac{4n}{d^2} - \tfrac{6n}{d^3} + 2nd^2 + \tfrac{d^5}{5} + \tfrac{d^3}{3} + \tfrac{3n^2}{d^3} + \tfrac{d^4}{2} + nd^3 + 2n^2 d + nd - \tfrac{5n^2}{d} + \tfrac{4n}{d} \\
&\leq 2n + 2nd^2 + \tfrac{d^5}{5} + \tfrac{d^3}{3} + \tfrac{3n^2}{d^3} + \tfrac{d^4}{2} + nd^3 + 2n^2 d + nd + \tfrac{4n}{d} \\
&\leq 6n^2 d,
\end{aligned}$$

where in the last step we have used only trivial bounds involving the facts that $n \geq d^2$ and $d \geq 2$.

For a fixed λ, let $\mathcal{L}(\lambda) \coloneqq \{i \in [d] \mid |L_i| \geq 5\sqrt{n}\}$. Then

$$\mathop{\mathbf{E}}_{\boldsymbol{\lambda} \sim \mathrm{SW}^n_d} \left[\sum_{i \in \mathcal{L}(\boldsymbol{\lambda})} \boldsymbol{L}_i^2 \right] \leq \frac{1}{25n} \mathop{\mathbf{E}}_{\boldsymbol{\lambda} \sim \mathrm{SW}^n_d} \left[\sum_{i \in \mathcal{L}(\boldsymbol{\lambda})} \boldsymbol{L}_i^4 \right] \leq \frac{1}{25n} \mathop{\mathbf{E}}_{\boldsymbol{\lambda} \sim \mathrm{SW}^n_d} \left[\sum_{i=1}^d \boldsymbol{L}_i^4 \right] \leq \frac{nd}{4}.$$

Thus, by (15),

$$\mathop{\mathbf{E}}_{\boldsymbol{\lambda} \sim \mathrm{SW}^n_d} \left[\sum_{i \in [d] \setminus \mathcal{L}(\boldsymbol{\lambda})} \boldsymbol{L}_i^2 \right] \geq \frac{nd}{2}.$$

Now define

$$\mathcal{M}(\lambda) \coloneqq \left\{ i \in [d] \;\middle|\; \frac{\sqrt{n}}{200} \leq |L_i| < 5\sqrt{n} \right\},$$

and let $\mathcal{E}$ be the event that $|\mathcal{M}(\lambda)| \geq d/200$. We claim that $p = \mathbf{Pr}[\mathcal{E}] \geq 1/100$. This is because if $p < 1/100$, then

$$\mathop{\mathbf{E}}_{\boldsymbol{\lambda} \sim \mathrm{SW}^n_d} \left[\sum_{i \in [d] \setminus \mathcal{L}(\boldsymbol{\lambda})} \boldsymbol{L}_i^2 \right] \leq p \cdot 25nd + (1-p) \cdot \left(\frac{25nd}{200} + \left(1 - \frac{1}{200}\right) \cdot \frac{nd}{200^2} \right) < \frac{nd}{2},$$

which is a contradiction.

Now let us use the assumption that $n \geq 10^{10} d^2$. Consider any coordinate $i \in [d]$ satisfying

$$|\boldsymbol{L}_i| = \left| \boldsymbol{\lambda}_i - i - \frac{n}{d} \right| \geq \frac{\sqrt{n}}{200}.$$

By our assumption that $n \geq 10^{10} d^2$, this implies that

$$\left| \boldsymbol{\lambda}_i - \frac{n}{d} \right| \geq \frac{\sqrt{n}}{1000}.$$

As a result, when $\mathcal{E}$ holds, which happens with at least 1% probability, there are at least $\frac{d}{200}$ coordinates $i \in [d]$ such that

$$\left| \boldsymbol{\lambda}_i - \frac{n}{d} \right| \geq \frac{\sqrt{n}}{1000}.$$

This completes the proof. □

4 A quantum Paninski theorem

In this section, we prove Theorem 1.10, that $\Theta(d/\epsilon^2)$ copies are necessary and sufficient to test whether or not a given state $\rho \in \mathbb{C}^{d\times d}$ is the maximally mixed state, i.e., has spectrum $(\frac{1}{d}, \ldots, \frac{1}{d})$.

4.1 The upper bound

The upper bound for Theorem 1.10 will follow from our analysis of the following simple algorithm.

Mixedness Tester. Given $\rho^{\otimes n}$, where ρ is d-dimensional:

1. Sample $\boldsymbol{\lambda} \sim \mathrm{SW}_\rho^n$.
2. Accept if $p_2^\sharp(\boldsymbol{\lambda}) \le \left(1 + \frac{\epsilon^2}{2}\right) \cdot \frac{n(n-1)}{d}$. Reject otherwise.

We remark that the tester Childs et al. [CHW07] used to distinguish the maximally mixed states of dimension $\frac{d}{2}$ and d also depended only on the magnitude of $p_2^\sharp(\boldsymbol{\lambda}) = 2c_1(\boldsymbol{\lambda})$; see [CHW07, equations (49), (50)].

Theorem 4.1. *The Mixedness Tester can test whether a state $\rho \in \mathbb{C}^{d\times d}$ is the maximally mixed state using $n = O(d/\epsilon^2)$ copies of ρ.*

Proof. We will run the Mixedness Tester with $n = 100d/\epsilon^2$. Both the "completeness" and the "soundness" analysis will require the last identity from (13), namely

$$(p_2^\sharp)^2 = p_{(2,2)}^\sharp + 4p_3^\sharp + 2p_{(1,1)}^\sharp. \tag{16}$$

Completeness. Suppose first that ρ is the maximally mixed state, so that in fact $\boldsymbol{\lambda} \sim \mathrm{SW}_d^n$. We compute the mean and variance of $p_2^\sharp(\boldsymbol{\lambda})$ using (16) and Corollary 2.35:

$$\mathop{\mathbf{E}}_{\boldsymbol{\lambda}\sim \mathrm{SW}_d^n}[p_2^\sharp(\boldsymbol{\lambda})] = \frac{n(n-1)}{d}, \tag{17}$$

$$\mathop{\mathbf{Var}}_{\boldsymbol{\lambda}\sim \mathrm{SW}_d^n}\left[p_2^\sharp(\boldsymbol{\lambda})\right] = \mathop{\mathbf{E}}_{\boldsymbol{\lambda}\sim \mathrm{SW}_d^n}\left[p_2^\sharp(\boldsymbol{\lambda})^2\right] - \left(\mathop{\mathbf{E}}_{\boldsymbol{\lambda}\sim \mathrm{SW}_d^n}[p_2^\sharp(\boldsymbol{\lambda})]\right)^2 = \frac{2n(n-1)(d^2-1)}{d^2} \le 2n(n-1). \tag{18}$$

Thus by Chebyshev's inequality,

$$\mathop{\mathbf{Pr}}_{\boldsymbol{\lambda}\sim \mathrm{SW}_d^n}\left[p_2^\sharp(\boldsymbol{\lambda}) > \left(1 + \frac{\epsilon^2}{2}\right) \cdot \frac{n(n-1)}{d}\right] \le \frac{8d^2}{n(n-1)\epsilon^4} \le \frac{1}{3},$$

by our choice of n. Thus indeed when ρ is the maximally mixed state, the Mixedness Tester accepts with probability at least $2/3$.

Soundness. Suppose now that ρ is a density matrix whose spectrum $\eta = (\eta_1, \ldots, \eta_d)$ satisfies $d_{\mathrm{TV}}^{\mathrm{sym}}(\eta, \mathsf{Unif}_d) \ge \epsilon$. Writing $\eta_i = \frac{1}{d} + \Delta_i$, this means that

$$\epsilon \le \frac{1}{2} \cdot \sum_{i=1}^d |\Delta_i| \le \frac{1}{2}\sqrt{d \cdot \sum_{i=1}^d \Delta_i^2},$$

using Cauchy–Schwarz; hence

$$\sum_{i=1}^{d} \Delta_i^2 \geq \frac{4\epsilon^2}{d}. \tag{19}$$

Using (16) and Proposition 2.34, we can calculate the difference between the mean of $p_2^\sharp(\boldsymbol{\lambda})$ and the cutoff used by the Mixedness Tester as

$$\begin{aligned}
\mathop{\mathbf{E}}_{\boldsymbol{\lambda}\sim \mathrm{SW}_\rho^n}\left[p_2^\sharp(\boldsymbol{\lambda})\right] - \frac{n(n-1)}{d}\cdot\left(1+\frac{\epsilon^2}{2}\right) &= n(n-1)\left(\sum_{i=1}^{d}\eta_i^2 - \frac{1}{d}\left(1+\frac{\epsilon^2}{2}\right)\right). \\
&= n(n-1)\left(\sum_{i=1}^{d}\Delta_i^2 - \frac{\epsilon^2}{2d}\right) \\
&\geq \frac{n(n-1)}{2}\sum_{i=1}^{d}\Delta_i^2,
\end{aligned}$$

where the last line follows from (19). Similarly, we can calculate the variance of $p_2^\sharp(\boldsymbol{\lambda})$ as

$$\begin{aligned}
\mathop{\mathbf{Var}}_{\boldsymbol{\lambda}\sim \mathrm{SW}_\rho^n}\left[p_2^\sharp(\boldsymbol{\lambda})\right] &= n(n-1)\left(2+4n\left(\sum\eta_i^3 - \left(\sum\eta_i^2\right)^2\right) + 6\left(\sum\eta_i^2\right)^2 - 8\sum\eta_i^3\right) \\
&\leq n(n-1)\left(8+4n\left(\sum\eta_i^3 - \left(\sum\eta_i^2\right)^2\right)\right) \\
&= n(n-1)\left(8+4n\left(\frac{1}{d}\sum\Delta_i^2 + \sum\Delta_i^3 - \left(\sum\Delta_i^2\right)^2\right)\right) \\
&\leq n(n-1)\left(8+8n\left(\sum\Delta_i^2\right)\right).
\end{aligned}$$

Applying Chebyshev's inequality gives us

$$\begin{aligned}
\mathop{\mathbf{Pr}}_{\boldsymbol{\lambda}\sim \mathrm{SW}_\rho^n}\left[p_{(2)}^\sharp(\boldsymbol{\lambda}) < \left(1+\frac{\epsilon^2}{2}\right)\cdot\frac{n(n-1)}{d}\right] &\leq \frac{1}{n(n-1)\left(\sum_{i=1}^{d}\Delta_i^2\right)^2}\cdot\left(32+32n\left(\sum_{i=1}^{d}\Delta_i^2\right)\right) \\
&\leq \frac{4}{n^2\left(\epsilon^2/d\right)^2} + \frac{16}{n\left(\epsilon^2/d\right)},
\end{aligned}$$

where the second step follows from (19). By our choice of n, this is at most $1/3$. Thus, when ρ is ϵ-far from the maximally mixed state, the Mixedness Tester rejects with probability at least $2/3$, as required. □

4.2 The lower bound: overview

For almost all of the lower bound proof we will assume d is even. In the end we will indicate how to obtain the lower bound when d is odd. For $0 \leq \epsilon \leq \frac{1}{2}$, let P_d^ϵ denote the probability distribution on $[d]$ in which

$$\mathsf{P}_d^\epsilon(j) = \frac{1+(-1)^{j-1}2\epsilon}{d}.$$

This is essentially the same probability distribution that Paninski [Pan08] studies in his lower bound. As usual, we also identify P_d^ϵ with the diagonal density matrix having these entries; i.e.,

$$\mathsf{P}_d^\epsilon = \operatorname{diag}\left(\frac{1+2\epsilon}{d}, \frac{1-2\epsilon}{d}, \frac{1+2\epsilon}{d}, \frac{1-2\epsilon}{d}, \ldots, \frac{1+2\epsilon}{d}, \frac{1-2\epsilon}{d}\right).$$

Note that $d_{\mathrm{TV}}^{\mathrm{sym}}(\mathsf{P}_d^\epsilon, \mathsf{Unif}_d) = \epsilon$. We also remark that when $\epsilon = \frac{1}{2}$, the distribution P_d^ϵ is the uniform distribution on $\frac{d}{2}$ elements (the odd-numbered ones). As in [Pan08], it proves to be most convenient to study the chi-squared distance between $\mathrm{SW}_{\mathsf{P}_d^\epsilon}^n$ and SW_d^n; our main theorem is the following:

Theorem 4.2. $d_{\chi^2}(\mathrm{SW}_{\mathsf{P}_d^\epsilon}^n, \mathrm{SW}_d^n) \le \exp((4n\epsilon^2/d)^2) - 1$.

Since this distance is small unless $n = \Omega(d/\epsilon^2)$, our lower bound is complete. More precisely:

Corollary 4.3. *For even d, testing whether a d-dimensional mixed state ρ has the the property of being the maximally mixed requires $n \ge .15d/\epsilon^2$ copies.*

Proof. In light of Lemma 2.22 we know that any ϵ-tester may as well make its testing decision based on a draw $\boldsymbol{\lambda} \sim \mathrm{SW}_\rho^n$. Since $d_{\mathrm{TV}}^{\mathrm{sym}}(\mathsf{P}_d^\epsilon, \mathsf{Unif}_d) = \epsilon$, the tester must be able to distinguish a draw from $\mathrm{SW}_{\mathsf{P}_d^\epsilon}^n$ and a draw from SW_d^n with probability advantage $1/3$; this is possible if and only if $d_{\mathrm{TV}}(\mathrm{SW}_{\mathsf{P}_d^\epsilon}^n, \mathrm{SW}_d^n) \ge 1/3$. But

$$d_{\mathrm{TV}}(\mathrm{SW}_{\mathsf{P}_d^\epsilon}^n, \mathrm{SW}_d^n) \le \frac{1}{2}\sqrt{d_{\chi^2}(\mathrm{SW}_{\mathsf{P}_d^\epsilon}^n, \mathrm{SW}_d^n)} \le \frac{1}{2}\sqrt{\exp((4n\epsilon^2/d)^2) - 1} < 1/3.$$

if $n < .15d/\epsilon^2$. □

We remark that by taking $\epsilon = \frac{1}{2}$ we exactly recover the lower bound from Theorem 1.9 due to Childs et al. [CHW07].

There are two major steps in the proof of Theorem 4.2. The first major step will be proving the following formula:

Theorem 4.4. *Let $x \in \mathbb{R}^d$ satisfy $x_1 + \cdots + x_d = 0$. Then*

$$\mathop{\mathbf{E}}_{\boldsymbol{\lambda} \sim \mathrm{SW}_d^n}\left[\left(\frac{s_{\boldsymbol{\lambda}}(1+x_1, \ldots, 1+x_d)}{s_{\boldsymbol{\lambda}}(1, \ldots, 1)} - 1\right)^2\right] = \sum_{\substack{\mu \in \mathrm{Par} \\ 0 < \ell(\mu) \le d}} \frac{s_\mu(x)^2}{d^{\uparrow \mu} \cdot d^{|\mu|}} \cdot n^{\downarrow |\mu|}.$$

(The sum has only finitely many terms since $n^{\downarrow |\mu|} = 0$ when $|\mu| > n$.)

Once the above theorem is established, the following consequence is essentially immediate:

Corollary 4.5. *Let $x \in \mathbb{R}^d$ satisfy $x_1 + \cdots + x_d = 0$ and $x_i \ge -1$ for all i. We write $\mathcal{Q}_x$ for the probability distribution on $[d]$ in which i has probability $\frac{1+x_i}{d}$. Then*

$$d_{\chi^2}(\mathrm{SW}_{\mathcal{Q}_x}^n, \mathrm{SW}_d^n) = \sum_{\substack{\mu \in \mathrm{Par} \\ 0 < \ell(\mu) \le d}} \frac{s_\mu(x)^2}{d^{\uparrow \mu} \cdot d^{|\mu|}} \cdot n^{\downarrow |\mu|}.$$

Proof. By definition, $d_{\chi^2}(\mathrm{SW}_{\mathcal{Q}_x}^n, \mathrm{SW}_d^n)$ is equal to

$$\mathop{\mathbf{E}}_{\boldsymbol{\lambda} \sim \mathrm{SW}_d^n}\left[\left(\frac{\mathrm{SW}_{\mathcal{Q}_x}^n(\boldsymbol{\lambda})}{\mathrm{SW}_d^n(\boldsymbol{\lambda})} - 1\right)^2\right] = \mathop{\mathbf{E}}_{\boldsymbol{\lambda} \sim \mathrm{SW}_d^n}\left[\left(\frac{s_{\boldsymbol{\lambda}}(\frac{1+x_1}{d}, \ldots, \frac{1+x_d}{d}) \dim(\boldsymbol{\lambda})}{s_{\boldsymbol{\lambda}}(\frac{1}{d}, \ldots, \frac{1}{d}) \dim(\boldsymbol{\lambda})} - 1\right)^2\right].$$

where we used Proposition 2.23. In turn, this equals the quantity on the left in Theorem 4.4 after canceling the common factor of $d^{-|\boldsymbol{\lambda}|} \dim \boldsymbol{\lambda}$ in the fraction (recall the homogeneity of the Schur polynomials). □

Let us sketch the intuition of the proof once Theorem 4.4 is established. We are ultimately interested in the case $x = 2\epsilon \cdot c$, where $\epsilon > 0$ is thought of as "small" and $c \in \mathbb{R}^d$ satisfies $c_1 + \cdots + c_d = 0$; specifically, $c = c_\pm := (+1, -1, +1, -1, \ldots, +1, -1)$. For simplicity, let us write ϵ instead of 2ϵ. Since s_μ is homogeneous of degree $|\mu|$, this means $s_\mu(x)^2 = s_\mu(c)^2 \epsilon^{2|\mu|}$. For the sake of intuition, let us consider the summands in Theorem 4.4 when $|\mu| = k$ is "small"; i.e., the coefficients on ϵ^{2k}. For $k = 1$ we have only $\mu = (1)$, and the associated summand actually drops out: this is because $s_{(1)}(x) = x_1 + \cdots + x_d = 0$. For $k \geq 2$, the term $n^{\downarrow |\mu|}$ is asymptotically n^k and the denominator $d^{|\mu|} \cdot d^{\uparrow \mu}$ is asymptotically d^{2k}. It remains to analyze $s_\mu(c_\pm)$. This is the second major step in the proof of Theorem 4.2: in Section 4.4 we establish an exact formula for it. Naively one might expect $|s_\mu(c_\pm)|$ to scale like d^k when $|\mu| = k$; however, as we will see it scales only like $d^{k/2}$ (and will in fact be 0 whenever k is odd). Thus the summands with $|\mu| = k$ small scale asymptotically as $n^k \cdot \frac{\epsilon^{2k}}{d^k}$, whence we get that $d_{\chi^2}(\mathrm{SW}^n_{\mathcal{Q}_{\epsilon \cdot c_\pm}}, \mathrm{SW}^n_d)$ is small if $n \ll \frac{d}{\epsilon^2}$.

4.3 Proof of Theorem 4.4

To analyze the quantity in Theorem 4.4 we will require the so-called Binomial Formula. (It generalizes the "usual" Binomial Formula, viz. $(1+x)^\ell = \sum_{m \geq 0} x^m \ell^{\downarrow m}/m!$, in the case $d = 1$.)

Theorem 4.6. *The following polynomial identity holds:*

$$\frac{s_\lambda(1 + x_1, \ldots, 1 + x_d)}{s_\lambda(1, \ldots, 1)} = \sum_{\substack{\mu \in \mathrm{Par} \\ \ell(\mu) \leq d}} \frac{s_\mu(x)}{d^{\uparrow \mu}} \cdot s^*_\mu(\lambda).$$

(The sum is actually finite since we may include the restriction $\mu \subseteq \lambda$ due to the factor $s^*_\mu(\lambda)$.)

In this form with the shifted Schur polynomials, the result appears in Okounkov and Olshanski's work [OO98b, Theorem 5.1] (see also [OO98a]). In a form involving factorial Schur polynomials it dates back to Lascoux [Las78]; see [Mac95, Example I.3.10].

The $\mu = \emptyset$ summand in Theorem 4.6 is always equal to 1; it follows that the quantity on the left of Theorem 4.4 is

$$\mathop{\mathbf{E}}_{\boldsymbol{\lambda} \sim \mathrm{SW}^n_d} \left[\left(\sum_{0 < \ell(\mu) \leq d} \frac{s_\mu(x)}{d^{\uparrow \mu}} \cdot s^*_\mu(\boldsymbol{\lambda}) \right)^2 \right] = \sum_{0 < \ell(\mu), \ell(\nu) \leq d} \frac{s_\mu(x) s_\nu(x)}{d^{\uparrow \mu} d^{\uparrow \nu}} \mathop{\mathbf{E}}_{\boldsymbol{\lambda} \sim \mathrm{SW}^n_d} \left[s^*_\mu(\boldsymbol{\lambda}) s^*_\nu(\boldsymbol{\lambda}) \right].$$

Therefore proving Theorem 4.4 reduces to proving

$$x_1 + \cdots + x_d = 0 \implies \sum_{0 < \ell(\mu), \ell(\nu) \leq d} \frac{s_\mu(x) s_\nu(x)}{d^{\uparrow \mu} d^{\uparrow \nu}} \mathop{\mathbf{E}}_{\boldsymbol{\lambda} \sim \mathrm{SW}^n_d} \left[s^*_\mu(\boldsymbol{\lambda}) s^*_\nu(\boldsymbol{\lambda}) \right] = \sum_{0 < \ell(\mu) \leq d} \frac{s_\mu(x)^2}{d^{\uparrow \mu} \cdot d^{|\mu|}} \cdot n^{\downarrow |\mu|}. \quad (20)$$

This is the main difficult step of the proof; the surprising aspect here is that we only get a contribution on the order of n^k from the terms with $|\mu| = k$, whereas naively one would expect n^{2k}. Showing that the $n^{k+1}, n^{k+2}, \ldots, n^{2k}$ contributions "drop out" is the essence of the proof.

In aid of proving (20), it's tempting to guess that $\mathbf{E}[s^*_\mu(\boldsymbol{\lambda}) s^*_\nu(\boldsymbol{\lambda})] = 1_{\{\mu = \nu\}} \cdot \frac{d^{\uparrow \mu}}{d^{|\mu|}} \cdot n^{\downarrow |\mu|}$; however such a statement is false. Instead, what *is* true is the following:

Theorem 4.7. *Let $x \in \mathbb{R}^d$ satisfy $x_1 + \cdots + x_d = 0$ and let $\mu \in \mathrm{Par}$ satisfy $|\mu| = r_1$ and $0 < \ell(\mu) \leq d$. Assume $r_2 \geq r_1$. Then*

$$\sum_{\substack{|\nu| = r_2 \\ \ell(\nu) \leq d}} \frac{s_\nu(x)}{d^{\uparrow \nu}} \mathop{\mathbf{E}}_{\boldsymbol{\lambda} \sim \mathrm{SW}^n_d} \left[s^*_\mu(\boldsymbol{\lambda}) s^*_\nu(\boldsymbol{\lambda}) \right] = 1_{\{r_2 = r_1\}} \cdot \frac{s_\mu(x)}{d^{|\mu|}} \cdot n^{\downarrow |\mu|}.$$

To deduce (20) from Theorem 4.7, simply write

$$\sum_{0<\ell(\mu),\ell(\nu)\le d} \frac{s_\mu(x)s_\nu(x)}{d^{\uparrow\mu}d^{\uparrow\nu}} \mathop{\mathbf{E}}_{\boldsymbol{\lambda}\sim \mathrm{SW}_d^n}\left[s_\mu^*(\boldsymbol{\lambda})s_\nu^*(\boldsymbol{\lambda})\right] = \sum_{r_1,r_2>0} \sum_{\substack{|\mu|=r_1\\ \ell(\mu)\le d}} \sum_{\substack{|\nu|=r_2\\ \ell(\nu)\le d}} \frac{s_\mu(x)s_\nu(x)}{d^{\uparrow\mu}d^{\uparrow\nu}} \mathop{\mathbf{E}}_{\boldsymbol{\lambda}\sim \mathrm{SW}_d^n}\left[s_\mu^*(\boldsymbol{\lambda})s_\nu^*(\boldsymbol{\lambda})\right].$$

Then use Theorem 4.7 when $r_2 \ge r_1$ and use it with the roles of μ and ν reversed when $r_2 < r_1$.

As for the proof of Theorem 4.7 itself, the first step is to compute the expected product of the shifted Schur polynomials. One possible approach for this might be to use the Littlewood–Richardson rule for factorial Schur functions (see [MS99, Proposition 4.2] or [Mol09, Corollary 3.3]) to write $s_\mu^* s_\nu^*$ as a linear combination of s_τ^* polynomials. Unfortunately, these Littlewood–Richardson coefficients seem somewhat difficult to work with. Instead, we will expand the shifted Schur polynomials in terms of the central characters and then multiply them via the known structure constants. We do this in the below lemma, carried out for a generic Schur–Weyl distribution. In this lemma, $\mathfrak{S}(R)$ denotes the symmetric group acting on the finite set R.

Lemma 4.8. *Let $q = (q_1, \dots, q_d)$ be a probability distribution on $[d]$ and let $\mu \vdash r_1$, $\nu \vdash r_2$. Then*

$$\mathop{\mathbf{E}}_{\boldsymbol{\lambda}\sim \mathrm{SW}_q^n}\left[s_\mu^*(\boldsymbol{\lambda})s_\nu^*(\boldsymbol{\lambda})\right] = \sum_{t=r_1\vee r_2}^{r_1+r_2} C^t_{r_1 r_2} \cdot n^{\downarrow t} \cdot \mathop{\mathbf{E}}_{\substack{\boldsymbol{w}_1\sim\mathfrak{S}(R_1)\\ \boldsymbol{w}_2\sim\mathfrak{S}(R_2)}} \left[\chi_\mu(\boldsymbol{w}_1)\chi_\nu(\boldsymbol{w}_2)p_{\overline{\boldsymbol{w}}_1\overline{\boldsymbol{w}}_2}(q)\right].$$

Here, for each choice of t, we let R_1, R_2 denote (arbitrary but fixed) subsets of $[t]$ having cardinality r_1, r_2, respectively, with $R_1 \cup R_2 = [t]$. (E.g., $R_1 = \{1, \dots, r_1\}$, $R_2 = \{t - r_2 + 1, \dots, t\}$.) Also, $\overline{\boldsymbol{w}}_1$ denotes the extension of $\boldsymbol{w}_1$ to $\mathfrak{S}_t$ formed by letting $\overline{\boldsymbol{w}}_1$ fix each element of $[t] \setminus R_1$; similarly for $\overline{\boldsymbol{w}}_2$.

Proof. Recall the notation $\rho(w)$ from Section 2.3 used denote the cycle type of a permutation w. In this proof, we also use the following notation: We write $\boldsymbol{\rho} \sim \mathfrak{S}_r$ to denote that $\boldsymbol{\rho}$ is a random partition of r formed by first choosing $\boldsymbol{w} \sim \mathfrak{S}_r$ uniformly and then taking $\boldsymbol{\rho} = \rho(\boldsymbol{w})$.

Using Theorem 2.33 for the first equality below, and Corollary 2.37 for the third equality, we have

$$\begin{aligned}
&\mathop{\mathbf{E}}_{\boldsymbol{\lambda}\sim \mathrm{SW}_q^n}\left[s_\mu^*(\boldsymbol{\lambda})s_\nu^*(\boldsymbol{\lambda})\right] \\
&= \mathop{\mathbf{E}}_{\boldsymbol{\lambda}\sim \mathrm{SW}_q^n}\left[\mathop{\mathbf{E}}_{\boldsymbol{\rho}_1\sim\mathfrak{S}_{r_1}}\left[\chi_\mu(\boldsymbol{\rho}_1)\cdot p^\sharp_{\boldsymbol{\rho}_1}(\boldsymbol{\lambda})\right]\cdot \mathop{\mathbf{E}}_{\boldsymbol{\rho}_2\sim\mathfrak{S}_{r_2}}\left[\chi_\nu(\boldsymbol{\rho}_2)\cdot p^\sharp_{\boldsymbol{\rho}_2}(\boldsymbol{\lambda})\right]\right] \\
&= \mathop{\mathbf{E}}_{\substack{\boldsymbol{\rho}_1\sim\mathfrak{S}_{r_1}\\ \boldsymbol{\rho}_2\sim\mathfrak{S}_{r_2}}}\left[\chi_\mu(\boldsymbol{\rho}_1)\chi_\nu(\boldsymbol{\rho}_2)\cdot \mathop{\mathbf{E}}_{\boldsymbol{\lambda}\sim \mathrm{SW}_q^n}\left[p^\sharp_{\boldsymbol{\rho}_1}(\boldsymbol{\lambda})\cdot p^\sharp_{\boldsymbol{\rho}_2}(\boldsymbol{\lambda})\right]\right] \\
&= \mathop{\mathbf{E}}_{\substack{\boldsymbol{\rho}_1\sim\mathfrak{S}_{r_1}\\ \boldsymbol{\rho}_2\sim\mathfrak{S}_{r_2}}}\left[\chi_\mu(\boldsymbol{\rho}_1)\chi_\nu(\boldsymbol{\rho}_2)\cdot \mathop{\mathbf{E}}_{\boldsymbol{\lambda}\sim \mathrm{SW}_q^n}\left[\sum_{t=r_1\vee r_2}^{r_1+r_2}\sum_{\tau\vdash t} C^t_{r_1 r_2}\cdot \mathop{\mathbf{Pr}}_{\substack{\boldsymbol{w}_1\sim\mathfrak{S}(R_1)|\boldsymbol{\rho}_1\\ \boldsymbol{w}_2\sim\mathfrak{S}(R_2)|\boldsymbol{\rho}_2}}\left[\rho(\overline{\boldsymbol{w}}_1\overline{\boldsymbol{w}}_2)=\tau\right]\cdot p^\sharp_\tau(\boldsymbol{\lambda})\right]\right],
\end{aligned}$$

where here $\boldsymbol{w}_i$ is chosen to be a uniformly random permutation on R_i (as in the lemma's statement),

conditioned on having cycle type $\boldsymbol{\rho}_i$. By Proposition 2.34 the above equals

$$\begin{aligned}&\mathop{\mathbf{E}}_{\substack{\boldsymbol{\rho}_1\sim\mathfrak{S}_{r_1}\\ \boldsymbol{\rho}_2\sim\mathfrak{S}_{r_2}}}\left[\chi_\mu(\boldsymbol{\rho}_1)\chi_\nu(\boldsymbol{\rho}_2)\cdot\sum_{t=r_1\vee r_2}^{r_1+r_2}\sum_{\tau\vdash t}C^t_{r_1r_2}\cdot\mathop{\mathbf{Pr}}_{\substack{\boldsymbol{w}_1\sim\mathfrak{S}(R_1)|\boldsymbol{\rho}_1\\ \boldsymbol{w}_2\sim\mathfrak{S}(R_2)|\boldsymbol{\rho}_2}}[\rho(\overline{\boldsymbol{w}}_1\overline{\boldsymbol{w}}_2)=\tau]\cdot n^{\downarrow t}\cdot p_\tau(q)\right]\\ &=\sum_{t=r_1\vee r_2}^{r_1+r_2}C^t_{r_1r_2}\cdot n^{\downarrow t}\cdot\mathop{\mathbf{E}}_{\substack{\boldsymbol{\rho}_1\sim\mathfrak{S}_{r_1},\ \boldsymbol{\rho}_2\sim\mathfrak{S}_{r_2}\\ \boldsymbol{w}_1\sim\mathfrak{S}(R_1)|\boldsymbol{\rho}_1\\ \boldsymbol{w}_2\sim\mathfrak{S}(R_2)|\boldsymbol{\rho}_2}}\left[\chi_\mu(\boldsymbol{\rho}_1)\chi_\nu(\boldsymbol{\rho}_2)\cdot\sum_{\tau\vdash t}1_{\{\rho(\overline{\boldsymbol{w}}_1\overline{\boldsymbol{w}}_2)=\tau\}}\cdot p_\tau(q)\right]\end{aligned}$$

The summation on the inside here simply equals $p_{\rho(\overline{\boldsymbol{w}}_1\overline{\boldsymbol{w}}_2)}(q)$; we may also replace $\chi_\mu(\boldsymbol{\rho}_1)$ with $\chi_\mu(\boldsymbol{w}_1)$, and similarly for $\chi_\nu(\boldsymbol{\rho}_2)$. Thus to complete the proof it remains to show that $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ have the same distribution as in the statement of the lemma. But this is clear: if we first pick a random permutation of r_i symbols, then take its cycle type, then set $\boldsymbol{w}_i$ to be a random permutation of r_i symbols of this cycle type, this is the same as simply taking $\boldsymbol{w}_i$ to be a uniformly random permutation of r_i symbols. □

We will also require the following Fourier-theoretic lemma:

Lemma 4.9. *For $u\in\mathfrak{S}_r$, $\nu\vdash r$, and $d\in\mathbb{Z}^+$,*

$$\mathop{\mathbf{E}}_{\boldsymbol{w}\sim\mathfrak{S}_r}[\chi_\nu(\boldsymbol{w})\cdot d^{\ell(u\boldsymbol{w})}]=\frac{\chi_\nu(u)d^{\uparrow\nu}}{r!}.$$

Proof. Define the class function e on $\mathfrak{S}_r$ by

$$e(v)=p_v(\underbrace{1,\dots,1}_{d\text{ entries}})=d^{\ell(v)}.$$

Since $\chi_\nu(\boldsymbol{w})=\chi_\nu(\boldsymbol{w}^{-1})$ because χ_ν is a class function, the quantity on the left in the proposition's statement is

$$\begin{aligned}\mathop{\mathbf{E}}_{\boldsymbol{w}\sim\mathfrak{S}_r}[\chi_\nu(\boldsymbol{w}^{-1})\cdot d^{\ell(u\boldsymbol{w})}]&=\mathop{\mathbf{E}}_{\boldsymbol{v}\sim\mathfrak{S}_r}[\chi_\nu(\boldsymbol{v}^{-1}u)\cdot d^{\ell(\boldsymbol{v})}]=(e*\chi_\nu)(u)=\sum_{\mu\vdash r}\widetilde{e*\chi_\nu}(\mu)\chi_\mu(u)\\&=\sum_{\mu\vdash r}\tfrac{1}{\dim\mu}\widetilde{e}(\mu)\widetilde{\chi_\nu}(\mu)\chi_\mu(u)=\tfrac{1}{\dim\nu}\widetilde{e}(\nu)\chi_\nu(u)=\tfrac{1}{\dim\nu}s_\nu(1,\dots,1)\chi_\nu(u)=\frac{\chi_\nu(u)d^{\uparrow\nu}}{r!},\end{aligned}$$

the last equality being Proposition 2.11. □

We can now complete the proof of Theorem 4.7 (and therefore also Theorem 4.4):

Proof of Theorem 4.7. We will use Lemma 4.8 in the case of SW_d^n, i.e., $q=(\frac1d,\dots,\frac1d)$; in this case, for $\tau\vdash t$ we have $p_\tau(q)=d^{\ell(\tau)-t}$. We thereby obtain

$$\begin{aligned}&\sum_{\substack{|\nu|=r_2\\ \ell(\nu)\le d}}\frac{s_\nu(x)}{d^{\uparrow\nu}}\mathop{\mathbf{E}}_{\boldsymbol{\lambda}\sim\mathrm{SW}_d^n}\left[s^*_\mu(\boldsymbol{\lambda})s^*_\nu(\boldsymbol{\lambda})\right]\\&\qquad=\sum_{t=r_2}^{r_1+r_2}C^t_{r_1r_2}\cdot\frac{n^{\downarrow t}}{d^t}\cdot\mathop{\mathbf{E}}_{\boldsymbol{w}_1\sim\mathfrak{S}(R_1)}\Big[\chi_\mu(\boldsymbol{w}_1)\cdot\sum_{\substack{|\nu|=r_2\\ \ell(\nu)\le d}}\frac{s_\nu(x)}{d^{\uparrow\nu}}\cdot\mathop{\mathbf{E}}_{\boldsymbol{w}_2\sim\mathfrak{S}(R_2)}\Big[\chi_\nu(\boldsymbol{w}_2)d^{\ell(\overline{\boldsymbol{w}}_1\overline{\boldsymbol{w}}_2)}\Big]\Big].\end{aligned}\tag{21}$$

(Here we are using the convention $\ell(\overline{\boldsymbol{w}}_1\overline{\boldsymbol{w}}_2) = \ell(\rho(\overline{\boldsymbol{w}}_1\overline{\boldsymbol{w}}_2))$.) We now would like to analyze the number of cycles of $\overline{\boldsymbol{w}}_1\overline{\boldsymbol{w}}_2$ within $\mathfrak{S}_t$. In $\overline{\boldsymbol{w}}_1$'s cycle decomposition, there are some cycles that act *only* on elements of $R_1 \setminus R_2$. Let's write $\ell^{\setminus}(\boldsymbol{w}_1)$ for the number of such cycles, and let's define $\overline{\boldsymbol{w}}_1^{\cap} \in \mathfrak{S}_t$ to be $\overline{\boldsymbol{w}}_1$ with those cycles deleted. Thus

$$\ell(\overline{\boldsymbol{w}}_1\overline{\boldsymbol{w}}_2) = \ell^{\setminus}(\boldsymbol{w}_1) + \ell(\overline{\boldsymbol{w}}_1^{\cap} \cdot \overline{\boldsymbol{w}}_2).$$

Next, let $\boldsymbol{w}_1^{\perp}$ denote the permutation obtained by deleting every element of $R_1 \setminus R_2$ from the cycle decomposition of $\overline{\boldsymbol{w}}_1^{\cap}$. Though $\boldsymbol{w}_1^{\perp}$ acts only on $R_1 \cap R_2$, we will view it as an element of $\mathfrak{S}(R_2)$. Although we don't have $\boldsymbol{w}_1^{\perp} \cdot \boldsymbol{w}_2 = \overline{\boldsymbol{w}}_1^{\cap} \cdot \overline{\boldsymbol{w}}_2$, it's not too hard to see that

$$\ell(\overline{\boldsymbol{w}}_1^{\cap} \cdot \overline{\boldsymbol{w}}_2) = \ell(\boldsymbol{w}_1^{\perp} \cdot \boldsymbol{w}_2).$$

Thus we obtain

$$(21) = \sum_{t=r_2}^{r_1+r_2} C^t_{r_1 r_2} \cdot \frac{n^{\downarrow t}}{d^t} \cdot \mathop{\mathbf{E}}_{\boldsymbol{w}_1 \sim \mathfrak{S}(R_1)}\Big[\chi_\mu(\boldsymbol{w}_1) d^{\ell^{\setminus}(\boldsymbol{w}_1)} \cdot \sum_{\substack{|\nu|=r_2 \\ \ell(\nu)\le d}} \frac{s_\nu(x)}{d^{\uparrow \nu}} \cdot \mathop{\mathbf{E}}_{\boldsymbol{w}_2 \sim \mathfrak{S}(R_2)}\Big[\chi_\nu(\boldsymbol{w}_2) d^{\ell(\boldsymbol{w}_1^{\perp} \cdot \boldsymbol{w}_2)}\Big]\Big].$$

Applying Lemma 4.9, we deduce

$$(21) = \sum_{t=r_2}^{r_1+r_2} C^t_{r_1 r_2} \cdot \frac{n^{\downarrow t}}{d^t} \cdot \mathop{\mathbf{E}}_{\boldsymbol{w}_1 \sim \mathfrak{S}(R_1)}\Big[\chi_\mu(\boldsymbol{w}_1) d^{\ell^{\setminus}(\boldsymbol{w}_1)} \cdot \frac{1}{r_2!} \sum_{\substack{|\nu|=r_2 \\ \ell(\nu)\le d}} s_\nu(x) \chi_\nu(\boldsymbol{w}_1^{\perp})\Big].$$

Notice that we may extend the summation over ν to include $\ell(\nu) > d$ as well: since x has d coordinates, $s_\nu(x) = 0$ anyway when $\ell(\nu) > d$ by Proposition 2.12. Having done this, we replace $s_\nu(x)$ with $\mathbf{E}_{\boldsymbol{v} \sim \mathfrak{S}_{r_2}}[\chi_\nu(\boldsymbol{v}) p_{\boldsymbol{v}}(x)]$, obtaining

$$\begin{aligned}(21) &= \sum_{t=r_2}^{r_1+r_2} C^t_{r_1 r_2} \cdot \frac{n^{\downarrow t}}{d^t} \cdot \mathop{\mathbf{E}}_{\boldsymbol{w}_1 \sim \mathfrak{S}(R_1)}\Big[\chi_\mu(\boldsymbol{w}_1) d^{\ell^{\setminus}(\boldsymbol{w}_1)} \cdot \frac{1}{r_2!} \sum_{|\nu|=r_2} \mathop{\mathbf{E}}_{\boldsymbol{v} \sim \mathfrak{S}_{r_2}}[\chi_\nu(\boldsymbol{v}) \cdot p_{\boldsymbol{v}}(x)] \chi_\nu(\boldsymbol{w}_1^{\perp})\Big] \\ &= \sum_{t=r_2}^{r_1+r_2} \frac{C^t_{r_1 r_2}}{r_2!} \cdot \frac{n^{\downarrow t}}{d^t} \cdot \mathop{\mathbf{E}}_{\boldsymbol{w}_1 \sim \mathfrak{S}(R_1)}\Big[\chi_\mu(\boldsymbol{w}_1) d^{\ell^{\setminus}(\boldsymbol{w}_1)} \cdot \mathop{\mathbf{E}}_{\boldsymbol{v} \sim \mathfrak{S}_{r_2}}\Big[p_{\boldsymbol{v}}(x) \cdot \sum_{|\nu|=r_2} \chi_\nu(\boldsymbol{v}) \chi_\nu(\boldsymbol{w}_1^{\perp})\Big]\Big].\end{aligned}$$

We claim that the inner expectation is 0 in most cases. First, $p_{\boldsymbol{v}}(x)$ vanishes whenever $\boldsymbol{v}$ has a fixed point, since $p_1(x) = x_1 + \cdots + x_d = 0$ by assumption. Next, suppose that $\boldsymbol{v}$ has no fixed points. By the orthogonality relations of representation theory, the innermost sum vanishes unless $\boldsymbol{v}$ and $\boldsymbol{w}_1^{\perp}$ are conjugate. Since $\boldsymbol{w}_1^{\perp} \in \mathfrak{S}(R_2)$ acts only on $\mathbb{R}_1 \cap R_2$, it *must* have a fixed point (and therefore not be conjugate to $\boldsymbol{v}$) unless $R_2 \setminus R_1 = \emptyset$. Since $r_2 \ge r_1$, this can only happen if $|\mu| = r_1 = r_2 = t$. We conclude that the inner expectation can only be nonzero in case $|\mu| = r_1 = r_2 = t$. In this case we have $C^t_{r_1 r_2} = r_2!$ and $\ell^{\setminus}(\boldsymbol{w}_1) = 0$, whence

$$(21) = 1_{\{r_2 = r_1\}} \cdot \frac{n^{\downarrow r_1}}{d^{r_1}} \cdot \mathop{\mathbf{E}}_{\boldsymbol{w}_1 \sim \mathfrak{S}_{r_1}}\Big[\chi_\mu(\boldsymbol{w}_1) \cdot \mathop{\mathbf{E}}_{\boldsymbol{v} \sim \mathfrak{S}_{r_1}}\Big[p_{\boldsymbol{v}}(x) \cdot \sum_{|\nu|=r_1} \chi_\nu(\boldsymbol{v}) \chi_\nu(\boldsymbol{w}_1^{\perp})\Big]\Big].$$

Once again, the summation is 0 if $\boldsymbol{v}$ and $\boldsymbol{w}_1$ are not conjugate; otherwise it equals $z_{\rho(\boldsymbol{w}_1)}$. Further, having chosen $\boldsymbol{w}_1$, the probability that $\boldsymbol{v}$ is conjugate to $\boldsymbol{w}_1$ is precisely $z^{-1}_{\rho(\boldsymbol{w}_1)}$. Thus these factors cancel and we obtain

$$(21) = 1_{\{r_2 = r_1\}} \cdot \frac{n^{\downarrow r_1}}{d^{r_1}} \cdot \mathop{\mathbf{E}}_{\boldsymbol{w}_1 \sim \mathfrak{S}_{r_1}}[\chi_\mu(\boldsymbol{w}_1) \cdot p_{\boldsymbol{w}_1}(x)] = 1_{\{r_2 = r_1\}} \cdot \frac{n^{\downarrow r_1}}{d^{r_1}} \cdot s_\mu(x),$$

completing the proof. □

4.4 A formula for $s_\mu(+1, -1, +1, -1, \dots)$

For this formula we will need to recall the notion of the 2-quotient of a partition. This definition essentially encodes the ways in which a partition can be tiled by dominoes.

Definition 4.10. Given a partition μ, a 2-*hook* in $[\mu]$ is a hook of length 2; i.e., a domino whose removal from $[\mu]$ results in a valid Young diagram.

Definition 4.11. A partition μ is said to be *balanced* (or to have an *empty* 2-*core*) if $[\mu]$ can be reduced to the empty diagram by successive removal of 2-hooks.

Definition 4.12. Given a partition μ we write $[\mu]_{\text{even}}$ (respectively, $[\mu]_{\text{odd}}$) for the set of boxes $\square \in [\mu]$ with even (respectively, odd) content $c(\square)$.

Remark 4.13. It's obvious from Definition 4.11 that if $\mu \vdash k$ is balanced then $|[\mu]_{\text{even}}| = k/2$. In fact, the converse also holds (this follows from, e.g., [JK81, Theorem 2.7.41]).

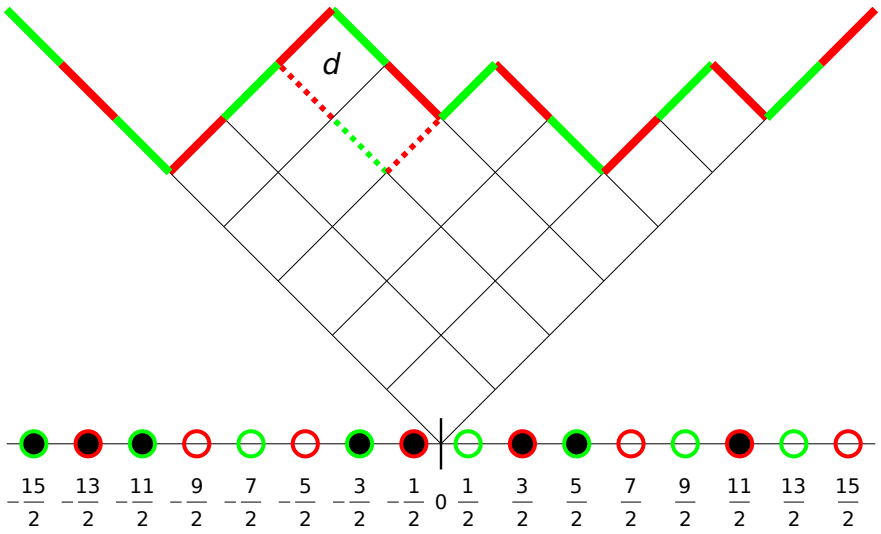


Figure 4: The Russian and Maya diagrams for $\mu = (6, 4, 4, 3, 3) \vdash 20$. The segments and pebbles corresponding to the 2-quotient pair are colored green and red. The dashed lines outline a 2-hook that could be removed; d is the square in this 2-hook with even content (namely, -2).

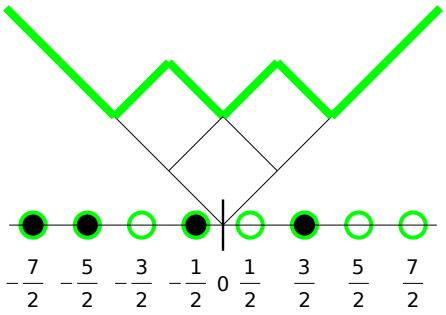


Figure 5: The diagram for 2-quotient partition $\mu^{(0)} = (2, 1) \vdash 3$.

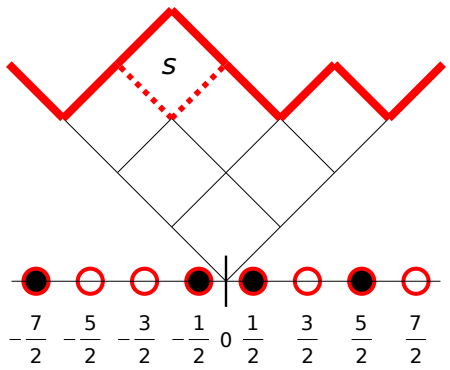


Figure 6: The diagram for 2-quotient partition $\mu^{(1)} = (3,2,2) \vdash 7$. The 1-hook square s (with content -1) is associated to the 2-hook in Figure 4 that contains square d.

Definition 4.14. Let μ be a partition. From the Maya diagram for $[\mu]$, form two new Maya diagrams by taking the two alternating sequences of pebbles. More precisely, for $b \in \{0,1\}$, let $\mu^{(b)}$ denote the partition whose Maya diagram is formed by the pebbles at positions $2z + (-1)^b \frac{1}{2}$, $z \in \mathbb{Z}$. (See Figure 4, in which $b = 0$ is associated to green and $b = 1$ is associated to red.) The pair $(\mu^{(0)}, \mu^{(1)})$ is called the 2-*quotient* of μ. (See Figures 5, 6 respectively.)

Remark 4.15. Note that when the Maya diagrams for $\mu^{(0)}$, $\mu^{(1)}$ are formed, each of the two origin mark positions may need to be adjusted from the former origin mark position coming from μ's origin mark. It is a fact (see, e.g., [RZ12, Section 2.1]) that μ is balanced if and only if *neither* origin mark position must be adjusted.

Fact 4.16. *A 2-hook in $[\mu]$ naturally corresponds to a sequence of three pebbles in $[\mu]$'s Maya diagram of the form (white, $*$, black). (See the dashed domino containing the label d in Figure 4.) In turn, this corresponds to a "1-hook" in one of $\mu^{(0)}, \mu^{(1)}$; i.e., a square on the rim whose removal leaves a valid Young diagram (see the square labeled s in Figure 6). Removal of the 2-hook from $[\mu]$ corresponds to replacing the sequence (white, $*$, black) by (black, $*$, white). (One thinks of the "filled" black pebble as jumping two positions to the left, onto the "empty" white pebble.) In turn, this corresponds to removing the associated 1-hook from either $\mu^{(0)}$ or $\mu^{(1)}$.*

We will require the following lemma. It is likely to be known; however we were unable to find its statement in the literature. The analogous lemma for hook lengths is well known (see, e.g., [RZ12, Lemma 2.1.ii]).

Lemma 4.17. *Let $\mu \vdash k$ be a balanced partition with 2-quotient $(\mu^{(0)}, \mu^{(1)})$. Then the multiset $\{c(\square) : \square \in [\mu^{(0)}], \square \in [\mu^{(1)}]\}$ is equal to the multiset $\{\frac{1}{2}c(\square) : \square \in [\mu]_{\text{even}}\}$.*

Proof. The statement is proved by induction on the deconstruction of μ from 2-hooks, with the base case being $\mu = \emptyset$. We rely on the fact that since μ is balanced, the Maya diagrams of $\mu^{(0)}$ and $\mu^{(1)}$ can be seen alternating within the Maya diagram for μ, with all three origin markers "lining up" (see Remark 4.15). By way of induction, suppose we consider the removal of some 2-hook D from $[\mu]$. This corresponds (see Fact 4.16) to removing a 1-hook (square) s from $\mu^{(b)}$, for some $b \in \{0,1\}$. Exactly one of D's two squares is in $[\mu]_{\text{even}}$; call that square d. (See Figures 4, 6 for illustration.) By induction, it suffices to show that $\frac{1}{2}c(d) = c(s)$. But this is easily seen from the combination of the Russian and Maya diagrams, as the content of a square is simply the horizontal displacement of its center. □

We are now ready to establish a formula for $s_\mu(+1,-1,+1,-1,\dots)$.

Theorem 4.18. *Let $\mu \vdash k$ and let d be even. Then*

$$s_\mu(\underbrace{+1,-1,+1,-1,\dots}_{d \text{ entries}}) = \begin{cases} 0 & \text{if } \mu \text{ is not balanced,} \\ \chi_\mu(\underbrace{2,2,\dots,2}_{k/2 \text{ entries}}) \cdot \frac{1}{k!!} \cdot (d^{\uparrow[\mu]_{\text{even}}}) & \text{if } \mu \text{ is balanced.} \end{cases}$$

Proof. The first part of the proof relies on a formula from [RSW04, Theorem 4.3], specialized to the case of "t" $= 2$:

$$s_\mu(\underbrace{+1,-1,+1,-1,\dots}_{d \text{ entries}}) = \begin{cases} 0 & \text{if } \mu \text{ is not balanced,} \\ \operatorname{sgn}(\chi_\mu(\underbrace{2,2,\dots,2}_{k/2 \text{ entries}})) \cdot s_{\mu^{(0)}}(\underbrace{1,1,\dots,1}_{d/2 \text{ entries}}) \cdot s_{\mu^{(1)}}(\underbrace{1,1,\dots,1}_{d/2 \text{ entries}}) & \text{if } \mu \text{ is balanced,} \end{cases}$$

where $(\mu^{(0)}, \mu^{(1)})$ is the 2-quotient of μ. Thus it suffices to show

$$s_{\mu^{(0)}}(1,1,\dots,1) \cdot s_{\mu^{(1)}}(1,1,\dots,1) = \frac{|\chi_\mu(2,2,\dots,2)| \cdot (d^{\uparrow[\mu]_{\text{even}}})}{(k/2)! \cdot 2^{k/2}} \tag{22}$$

assuming μ is balanced. Applying Proposition 2.11, the left-hand side of (22) is

$$\frac{(\frac{d}{2}^{\uparrow\mu^{(0)}}) \cdot (\frac{d}{2}^{\uparrow\mu^{(1)}}) \cdot \dim \mu^{(0)} \cdot \dim \mu^{(1)}}{|\mu^{(0)}|! \cdot |\mu^{(1)}|!}.$$

Next, we appeal to [RZ12, formula (2.2)], which states

$$\chi_\mu(2,2,\dots,2) = \sigma_\mu \cdot \binom{|\mu|/2}{|\mu^{(0)}|, |\mu^{(1)}|} \cdot \dim \mu^{(0)} \cdot \dim \mu^{(1)},$$

where $\sigma_\mu \in \{\pm 1\}$ is a certain sign. Thus to verify (22) it remains to show

$$(\tfrac{d}{2}^{\uparrow\mu^{(0)}}) \cdot (\tfrac{d}{2}^{\uparrow\mu^{(1)}}) = \frac{d^{\uparrow[\mu]_{\text{even}}}}{2^{k/2}}. \tag{23}$$

But this follows immediately from Lemma 4.17. □

4.5 Wrapping up the lower bound

In this section we complete the proof of Theorem 4.2. We begin by applying Corollary 4.5 with $x = (+2\epsilon, -2\epsilon, +2\epsilon, -2\epsilon, \dots)$. Using Theorem 4.18 and the homogeneity of Schur polynomials, we obtain the following after a few manipulations:

Theorem 4.19. *For d even and $0 \le \epsilon \le \frac{1}{2}$,*

$$d_{\chi^2}(\mathrm{SW}^n_{\mathbb{P}^\epsilon_d}, \mathrm{SW}^n_d) = \sum_{k=2,4,6,\dots} n^{\downarrow k} (2\epsilon)^{2k} d^{-k} \cdot \Big(\frac{1}{k!!^2} \sum_{\substack{\mu \vdash k \text{ balanced} \\ 0 < \ell(\mu) \le d}} \chi_\mu(2,\dots,2)^2 \cdot \frac{d^{\uparrow[\mu]_{\text{even}}}}{d^{\uparrow[\mu]_{\text{odd}}}} \Big). \tag{24}$$

To estimate this quantity we will use the following very crude bound:

Proposition 4.20. *Let $d \in \mathbb{Z}^+$ and let $\mu \vdash k$ be balanced, with $0 < \ell(\mu) \leq d$. Then*

$$\frac{d^{\uparrow [\mu]_{\text{even}}}}{d^{\uparrow [\mu]_{\text{odd}}}} \leq 2^{k/2}. \tag{25}$$

Proof. Fix any domino-tiling for μ. Each of the $k/2$ dominoes contains one cell of even content c_e and one cell of odd content c_o, with $|c_e - c_o| = 1$. Thus each contributes a factor of $\frac{d+c_e}{d+c_o} \leq \frac{2}{1} = 2$ to $(d^{\uparrow [\mu]_{\text{even}}})/(d^{\uparrow [\mu]_{\text{odd}}})$. □

By character orthogonality relations we also have

$$\sum_{\substack{\mu \vdash k \text{ balanced} \\ 0 < \ell(\mu) \leq d}} \chi_\mu(2, \dots, 2)^2 \leq \sum_{\mu \vdash k} \chi_\mu(2, \dots, 2)^2 = z_{(2,\dots,2)} = k!!. \tag{26}$$

Combining (25), (26), we get that the parenthesized expression in (24) is at most $2^{k/2}/k!! = 1/(k/2)!$. Using also $n^{\downarrow k} \leq n^k$, the right-hand side of (24) is thus bounded by

$$\sum_{k=2,4,6,\dots} n^k (2\epsilon)^{2k} d^{-k}/(k/2)! = \exp((4n\epsilon^2/d)^2) - 1,$$

completing the proof of Theorem 4.2.

We end by indicating how to obtain the testing lower bound in the case when $d \geq 3$ is odd. In this case we define P_d^ϵ to be $(\frac{1+2\epsilon}{d}, \frac{1-2\epsilon}{d}, \dots, \frac{1+2\epsilon}{d}, \frac{1-2\epsilon}{d}, \frac{1}{d})$. This distribution has $d_{\mathrm{TV}}^{\mathrm{sym}}(\mathsf{P}_d^\epsilon, \mathsf{Unif}_d) = \frac{d-1}{d}\epsilon \geq \frac{2}{3}\epsilon$; since this differs from ϵ only by a constant factor, the lower bound of $\Omega(d/\epsilon^2)$ is not affected. Now Corollary 4.5 is applied with $x = (+2\epsilon, -2\epsilon, \dots, +2\epsilon, -2\epsilon, 0)$. By stability of the shifted Schur polynomials we have $s_\mu(+1, -1, \dots, +1, -1, 0) = s_\mu(+1, -1, \dots, +1, -1)$, where there are $d-1$ entries in the latter. Now we get $\chi_\mu(2, 2, \dots, 2) \cdot \frac{1}{k!!} \cdot (d-1)^{\uparrow [\mu]_{\text{even}}}$ out of Theorem 4.18, and we can simply upper-bound $(d-1)$ by d and proceed with the remainder of the proof.

5 Hardness of distinguishing uniform distributions

In this section, we prove Theorem 1.12, namely that $O(r^2/\Delta)$ copies are sufficient to distinguish between the cases when ρ's spectrum is uniform on either r or $r + \Delta$ eigenvalues ($1 \leq \Delta \leq r$), and that $\widetilde{\Omega}(r^2/\Delta)$ copies are necessary. To be more precise, our lower bound on the number of copies n will be

$$n \geq r^{2 - O(1/\log^{.33} r)}/\Delta. \tag{27}$$

5.1 The upper bound

The proof of the upper bound is quite similar to that of Theorem 4.1 for the Mixedness Tester. We employ the following tester:

Uniform Distribution Distinguisher. Given $\rho^{\otimes n}$:

1. Sample $\boldsymbol{\lambda} \sim \mathrm{SW}_\rho^n$.
2. Accept if $p_2^\sharp(\boldsymbol{\lambda}) \leq e := n(n-1) \cdot \frac{1}{2}\left(\frac{1}{r} + \frac{1}{r+\Delta}\right)$. Reject otherwise.

As for the analysis, from Equations (17) and (18):

$$\mathop{\mathbf{E}}_{\boldsymbol{\lambda}\sim \mathrm{SW}_m^n}\left[p_2^\sharp(\boldsymbol{\lambda})\right] = \frac{n(n-1)}{m}, \quad \text{and} \quad \mathop{\mathbf{Var}}_{\boldsymbol{\lambda}\sim \mathrm{SW}_m^n}\left[p_2^\sharp(\boldsymbol{\lambda})\right] \le 2n(n-1).$$

We see that the variance is the same whether $m = r$ or $m = r+\Delta$; only the expectation is different, and the tester's acceptance cutoff e is precisely the midway point between the two expectations. If $m = r$, then Chebyshev's inequality implies

$$\mathop{\mathbf{Pr}}_{\boldsymbol{\lambda}\sim \mathrm{SW}_m^n}\left[p_2^\sharp(\boldsymbol{\lambda}) \ge e\right] \le \frac{8r^2(r+\Delta)^2}{n(n-1)\Delta^2} \le \frac{32r^4}{(n-1)^2\Delta^2},$$

and we have the same upper bound by Chebyshev for $\mathbf{Pr}_{\boldsymbol{\lambda}\sim \mathrm{SW}_m^n}\left[p_2^\sharp(\boldsymbol{\lambda}) \le e\right]$ when $m = r+\Delta$. This upper bound is at most $1/3$ provided $n \ge 4\sqrt{6}\cdot\frac{r^2}{\Delta}+1$, completing the proof of the upper bound in Theorem 1.12.

The end of Section 6.1 gives a different $O(r^2)$-copy tester (the "Rank Tester") for the r-versus-$(r+1)$ case. In this case it's superior to the Uniform Distribution Distinguisher in that it has one-sided error (i.e., it never rejects in the rank-r case).

5.2 The lower bound

The bulk our work for the lower bound will be devoted to the case of $\Delta = 1$. The extension to larger Δ is very tedious and will be dealt with in Section 5.3. So let $r \in \mathbb{Z}^+$ be a parameter which we think of as tending to infinity, and for brevity let $r_+ = r+1$. Our task is to show that the distributions SW_r^n and $\mathrm{SW}_{r_+}^n$ are very close in total variation distance unless $n \ge \widetilde{\Omega}(r^2)$. For notational convenience we will write

$$n = \frac{r^2}{\omega^2}$$

and seek to show that SW_r^n and $\mathrm{SW}_{r_+}^n$ are close once ω is sufficiently large as a function of r. Ultimately we will select $\omega = \exp(\Theta(\log^{.67} r))$. For now, though, let's keep ω general, subjecting it only to the following assumption:

$$200 \le \omega \le \sqrt{r}. \tag{28}$$

5.2.1 Initial approximations

It proves more convenient to study the Kullback–Leibler divergence between SW_r^n and $\mathrm{SW}_{r_+}^n$:

$$\begin{aligned} d_{\mathrm{KL}}(\mathrm{SW}_r^n, \mathrm{SW}_{r_+}^n) &= \mathop{\mathbf{E}}_{\boldsymbol{\lambda}\sim \mathrm{SW}_r^n}\left[\ln\left(\frac{\mathrm{SW}_r^n[\boldsymbol{\lambda}]}{\mathrm{SW}_{r_+}^n[\boldsymbol{\lambda}]}\right)\right] \\ &= \mathop{\mathbf{E}}_{\boldsymbol{\lambda}\sim \mathrm{SW}_r^n}\left[\ln\left(\frac{r_+^n}{r^n}\cdot\frac{r^{\uparrow\boldsymbol{\lambda}}}{r_+^{\uparrow\boldsymbol{\lambda}}}\right)\right] \\ &= n\ln\left(\frac{r_+}{r}\right) + \mathop{\mathbf{E}}_{\boldsymbol{\lambda}\sim \mathrm{SW}_r^n}\left[\ln\left(\frac{\prod_{\square\in[\boldsymbol{\lambda}]}(r+c(\square))}{\prod_{\square\in[\boldsymbol{\lambda}]}(r_++c(\square))}\right)\right], \end{aligned} \tag{29}$$

where the second equality used Proposition 2.26. (We remark that the logarithms above are always finite since $\mathrm{supp}(\mathrm{SW}_r^n) \subseteq \mathrm{supp}(\mathrm{SW}_{r_+}^n)$.)

Recalling that $r_+ = r + 1$, it is very easy to verify (cf. [Mac95, Exercise I.1.11], [CGS04, Section 2.5]) that the large fraction inside the inner logarithm of (29) is equal to

$$\prod_{i=1}^{\ell(\boldsymbol{\lambda})} \frac{r-(i-1)}{r-(i-1-\boldsymbol{\lambda}_i)} = \Phi(-(r+\tfrac{1}{2}); \boldsymbol{\lambda}),$$

where Φ denotes a generating function for the modified Frobenius coordinates, defined in [IO02] and similar to the "Frobenius function" from [Las08, CSST10]. Proposition 1.2 in [IO02] observes that

$$\Phi(z;\lambda) = \prod_i \frac{z+b_i^*}{z-a_i^*},$$

where the a_i^*'s and b_i^*'s are the modified Frobenius coordinates of λ; as a consequence, Proposition 1.4 in [IO02] states that

$$\ln \Phi(z;\lambda) = \sum_{k=1}^{\infty} \frac{p_k^*(\lambda)}{k} z^{-k}. \tag{30}$$

However we cannot immediately take $z = -(r+\frac{1}{2})$ and conclude

$$(29) \stackrel{?}{=} n \ln\left(1+\frac{1}{r}\right) + \mathop{\mathbf{E}}_{\boldsymbol{\lambda}\sim \mathrm{SW}_r^n}\left[\sum_{k=1}^{\infty} \frac{(-1)^k p_k^*(\boldsymbol{\lambda})}{k(r+\frac{1}{2})^k}\right] \tag{31}$$

because (30) is merely a formal identity of generating functions and does not hold for all real z. More specifically, it's necessary that the Taylor series for $\ln(1+b_i/z)$ and $\ln(1-a_i/z)$ converge, which happens provided $|b_i/(r+\frac{1}{2})|, |a_i/(r+\frac{1}{2})| \le 1$. These conditions are equivalent to $\ell(\boldsymbol{\lambda}) = \boldsymbol{\lambda}_1' \le r+1$ and $\boldsymbol{\lambda}_1 \le r+1$. The first condition is automatic, since $\boldsymbol{\lambda} \sim \mathrm{SW}_r^n$. The second condition does not always hold; however, we will show (see Lemma 5.2 below) that it holds with overwhelming probability when $n \ll r^2$. Indeed the "central limit theorems" for the Schur–Weyl distributions suggest that both $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_1'$ will almost always be $O(\sqrt{n}) = O(\frac{r}{\omega})$. Let us therefore make a definition:

Definition 5.1. We say that $\lambda \vdash n$ is *usual* if $\lambda_1, \lambda_1' \le \frac{10}{\omega} r$. Since we are assuming $\omega \ge 200$, usual λ's satisfy $\lambda_1, \lambda_1' \le \frac{1}{20} r \le r+1$.

Thus when λ is usual we may apply (31). Since the quantity inside the expectation in (29) is clearly always negative, we may write

$$\begin{aligned}
d_{\mathrm{KL}}(\mathrm{SW}_r^n, \mathrm{SW}_{r_+}^n) = (29) &\le n \ln\left(1+\frac{1}{r}\right) + \mathop{\mathbf{E}}_{\boldsymbol{\lambda}\sim \mathrm{SW}_r^n}\left[1_{\{\boldsymbol{\lambda} \text{ usual}\}} \cdot \ln\left(\frac{\prod_{\square\in[\boldsymbol{\lambda}]}(r+c(\square))}{\prod_{\square\in[\boldsymbol{\lambda}]}(r_+ + c(\square))}\right)\right] \\
&= n \ln\left(1+\frac{1}{r}\right) + \mathop{\mathbf{E}}_{\boldsymbol{\lambda}\sim \mathrm{SW}_r^n}\left[1_{\{\boldsymbol{\lambda} \text{ usual}\}} \cdot \sum_{k=1}^{\infty} \frac{(-1)^k p_k^*(\boldsymbol{\lambda})}{k(r+\frac{1}{2})^k}\right] \\
&= n \ln\left(1+\frac{1}{r}\right) - \frac{1}{r+\frac{1}{2}} \cdot \mathop{\mathbf{E}}_{\boldsymbol{\lambda}\sim \mathrm{SW}_r^n}\left[1_{\{\boldsymbol{\lambda} \text{ usual}\}} \cdot p_1^*(\boldsymbol{\lambda})\right] \qquad (32) \\
&\quad + \mathop{\mathbf{E}}_{\boldsymbol{\lambda}\sim \mathrm{SW}_r^n}\left[1_{\{\boldsymbol{\lambda} \text{ usual}\}} \cdot \sum_{k=2}^{\infty} \frac{(-1)^k p_k^*(\boldsymbol{\lambda})}{k(r+\frac{1}{2})^k}\right]. \qquad (33)
\end{aligned}$$

Recall that $p_1^*(\lambda)$ is simply $|\lambda|$; thus the expectation in (32) is simply $n\,\mathbf{Pr}[\boldsymbol{\lambda}\text{ usual}]$. As Lemma 5.2 below shows, $\mathbf{Pr}[\boldsymbol{\lambda}\text{ usual}] = 1-\delta$ for $\delta \lll \frac{1}{60r^2}$. Thus:

$$(32) = n\left(\ln\left(1+\frac{1}{r}\right) - \frac{1}{r+\frac{1}{2}} + \frac{\delta}{r+\frac{1}{2}}\right) \le n\left(\frac{1}{12r^3} + \frac{1/(60r^2)}{r+\frac{1}{2}}\right) \le \frac{n}{10r^3} = \frac{1}{10\omega^2 r}. \tag{34}$$

Lemma 5.2. *Let $\boldsymbol{\lambda} \sim \mathrm{SW}_r^n$. Then $\mathbf{Pr}[\boldsymbol{\lambda}\text{ unusual}] \le 2^{-20r/\omega}$.*

Proof. Write $B = \lceil \frac{10}{\omega} r \rceil$. By Proposition 2.31 and the fact that $B \le r$,

$$\mathbf{Pr}[\boldsymbol{\lambda}_1 \ge B], \mathbf{Pr}[\boldsymbol{\lambda}_1' \ge B] \le \left(\frac{2e^2 n}{B^2}\right)^B \le \left(\frac{2e^2}{100}\right)^{10r/\omega} \le 2^{-1-20r/\omega}.$$

The lemma now follows from the union bound. □

Turning to (33), let's write

$$L_C^*(\lambda) \coloneqq \sum_{k=2}^{C} \frac{(-1)^k p_k^*(\lambda)}{k(r+\frac{1}{2})^k},$$

recalling that $L_\infty^*(\lambda)$ is definitely convergent if λ is usual. The infinite sum in (33) is inconvenient, as is the $+\frac{1}{2}$ in the denominator. We clean these issues up with the following lemma:

Lemma 5.3. *Assuming $\lambda \vdash n$ is usual, if*

$$C \ge \frac{3\log(10r)}{\log(\omega/10)},$$

it follows that

$$|L_\infty^*(\lambda) - L_C(\lambda)| \le \frac{201}{\omega^3},$$

where $L_C(\lambda)$ denotes the same quantity as $L_C^*(\lambda)$ except with no $+\frac{1}{2}$ in the denominator.

Proof. For any $\lambda \vdash n$ (not necessarily usual), we have the crude bound $|p_k^*(\lambda)| \le 2\sqrt{n}B^k$ whenever $\lambda_1, \lambda_1' \le B$. This is because each modified Frobenius coordinate a_i^* or b_i^* (of which there are at most $\sqrt{n}$ each) is at most B. For *usual* λ we may take $B = \frac{10}{\omega}r$. Thus we have

$$|L_\infty^*(\lambda) - L_C^*(\lambda)| \le \sum_{k=C+1}^{\infty} \frac{|p_k^*(\lambda)|}{k(r+\frac{1}{2})^k} \le \sum_{k=C+1}^{\infty} \frac{2\frac{r}{\omega}(10\frac{r}{\omega})^k}{kr^k} \le 2r \sum_{k=C+1}^{\infty} \left(\frac{10}{\omega}\right)^k \le 4r\left(\frac{10}{\omega}\right)^C \le \frac{1}{250r^2},$$

where the last inequality used the assumption about C (and the second-to-last inequality used $\omega \ge 200$ in a crude way). Further,

$$|L_C^*(\lambda) - L_C(\lambda)| \le \sum_{k=2}^{C} \frac{|p_k^*(\lambda)|}{k}\left(\frac{1}{r^k} - \frac{1}{(r+\frac{1}{2})^k}\right) \le \sum_{k=2}^{C} \frac{2\frac{r}{\omega}(10\frac{r}{\omega})^k}{k}\left(\frac{k}{2r^{k+1}}\right) = \frac{1}{\omega}\sum_{k=2}^{C}\left(\frac{10}{\omega}\right)^k \le \frac{200}{\omega^3}.$$

Finally, $\frac{200}{\omega^3} + \frac{1}{250r^2} \le \frac{201}{\omega^3}$ by our assumption (28) that $\omega \le \sqrt{r}$. □

Let us use this lemma in (33), and also apply (34) in (32). Assuming the lemma's hypotheses, we obtain

$$\begin{aligned} d_{\mathrm{KL}}(\mathrm{SW}_r^n, \mathrm{SW}_{r_+}^n) &\le \mathop{\mathbf{E}}_{\boldsymbol{\lambda}\sim\mathrm{SW}_r^n}\left[1_{\{\boldsymbol{\lambda} \text{ usual}\}} \cdot L_C(\boldsymbol{\lambda})\right] + \tfrac{1}{10\omega^2 r} + \tfrac{201}{\omega^3} \\ &\le \mathop{\mathbf{E}}_{\boldsymbol{\lambda}\sim\mathrm{SW}_r^n}\left[L_C(\boldsymbol{\lambda})\right] - \mathop{\mathbf{E}}_{\boldsymbol{\lambda}\sim\mathrm{SW}_r^n}\left[1_{\{\boldsymbol{\lambda} \text{ unusual}\}} \cdot L_C(\boldsymbol{\lambda})\right] + \tfrac{202}{\omega^3}. \end{aligned}$$

We can use Cauchy–Schwarz to bound

$$\left|\mathop{\mathbf{E}}_{\boldsymbol{\lambda}\sim\mathrm{SW}_r^n}\left[1_{\{\boldsymbol{\lambda} \text{ unusual}\}} \cdot L_C(\boldsymbol{\lambda})\right]\right| \le \sqrt{\mathbf{E}[1_{\{\boldsymbol{\lambda} \text{ unusual}\}}^2]}\sqrt{\mathbf{E}[L_C(\boldsymbol{\lambda})^2]} \le 2^{-10r/\omega}\sqrt{\mathbf{E}[L_C(\boldsymbol{\lambda})^2]}, \tag{35}$$

where the last inequality used Lemma 5.2. Finally, we can afford to use an extraordinarily crude bound on $\mathbf{E}[L_C(\boldsymbol{\lambda})^2]$:

$$\mathbf{E}[L_C(\boldsymbol{\lambda})^2] \le C\sum_{k=2}^{C}\mathbf{E}[p_k^*(\boldsymbol{\lambda})^2] \le C\sum_{k=2}^{C}(2\sqrt{n}n^k)^2 \le n^{3C} \le r^{6C},$$

where the second inequality used the crude bound on $|p_k^*(\lambda)|$ from the proof of Lemma 5.3. (In fact, in Section 5.3 we will actually show that this quantity is quite tiny.) If we now make the very weak assumption that $C \le \frac{3r}{\omega \log r}$, we may conclude (35) $\le 2^{-r/\omega} \ll \frac{1}{\omega^3}$.

Now we can summarize all of the preparatory work we have done so far:

Proposition 5.4. *Assuming* $\frac{3\log(10r)}{\log(\omega/10)} \le C \le \frac{3r}{\omega\log r}$, *for* $\boldsymbol{\lambda} \sim \mathrm{SW}_r^n$ *we have*

$$d_{\mathrm{KL}}(\mathrm{SW}_r^n, \mathrm{SW}_{r_+}^n) \le \mathbf{E}\left[L_C(\boldsymbol{\lambda})\right] + \tfrac{203}{\omega^3},$$

where

$$L_C(\lambda) \coloneqq \sum_{k=2}^{C}\frac{(-1)^k p_k^*(\lambda)}{kr^k}. \tag{36}$$

(It is straightforward to check using (28) that the range of values for C is nonempty.)

We now come to the main task: showing that $\mathbf{E}[L_C(\boldsymbol{\lambda})]$ is small.

5.2.2 Passing to the $p^\sharp$ polynomials

In this section and the following one, we will use the notation

$$\mathrm{fact}(\mu) = \prod_{w\ge 1} m_w(\mu)!$$

where, recall, $m_w(\mu)$ is the number of parts of μ equal to w.

The following proposition is essentially immediate from known formulas:

Proposition 5.5. *For any* $k \in \mathbb{Z}^+$, *we have the following identity on observables:*

$$p_k^* = \sum_{\mu\ :\ \mathrm{wt}(\mu)=k+1}\frac{k^{\downarrow(\ell(\mu)-1)}}{\mathrm{fact}(\mu)}p_\mu^\sharp + \mathcal{O}_k,$$

where $\mathcal{O}_k$ *is an observable with* $\mathrm{wt}(\mathcal{O}_k) \le k$. *More precisely,*

$$\mathcal{O}_k = \sum_{\mu\ :\ \mathrm{wt}(\mu)\le k} c_{k,\mu}p_\mu^\sharp$$

for some rational coefficients $c_{k,\mu}$.

Proof. From [IO02, Corollary 2.8] we have

$$p_k^* = \frac{1}{k+1} \cdot \widetilde{p}_{k+1} + \Big\{\text{a linear combination of } \widetilde{p}_k, \dots, \widetilde{p}_2\Big\}.$$

From [IO02, Corollary 3.7] (cf. [Mél10b, Lemma 10.10]) we have

$$\widetilde{p}_{k+1} = \sum_{\mu \,:\, \mathrm{wt}(\mu)=k+1} \frac{(k+1)^{\downarrow \ell(\mu)}}{\mathrm{fact}(\mu)} \prod_{i \geq 1} (p_i^\sharp)^{m_i(\mu)}.$$

The result is now easily deduced from Proposition 2.39. □

Substituting the above result into (36) yields:

$$L_C(\lambda) = \sum_{k=2}^{C} \frac{(-1)^k}{kr^k} \cdot \sum_{\mathrm{wt}(\mu)=k+1} \frac{k^{\downarrow(\ell(\mu)-1)}}{\mathrm{fact}(\mu)} p_\mu^\sharp(\lambda) + \sum_{k=2}^{C} \frac{(-1)^k \mathcal{O}_k(\lambda)}{kr^k}. \tag{37}$$

Taking the expectation over $\boldsymbol{\lambda} \sim \mathrm{SW}_r^n$, and using Corollary 2.35 to evaluate the expectation of $p_\mu^\sharp$, we obtain:

$$\mathop{\mathbf{E}}_{\boldsymbol{\lambda} \sim \mathrm{SW}_r^n}[L_C(\boldsymbol{\lambda})] = \sum_{k=2}^{C} \frac{(-1)^k}{kr^k} \cdot \sum_{\mathrm{wt}(\mu)=k+1} \frac{k^{\downarrow(\ell(\mu)-1)}}{\mathrm{fact}(\mu)} n^{\downarrow|\mu|} r^{\ell(\mu)-|\mu|} \tag{38}$$

$$+ \sum_{k=2}^{C} \frac{(-1)^k \mathbf{E}_{\boldsymbol{\lambda} \sim \mathrm{SW}_r^n}[\mathcal{O}_k(\boldsymbol{\lambda})]}{kr^k}. \tag{39}$$

We will show in Lemmas 5.7, 5.8 below that the "error term" (39) is small assuming $n \ll r^2$. Thus we focus on (38).

5.2.3 Showing the "main term" is small: some intuition

Before diving into manipulations, let's take a high-level look at the contributions to (38) from $k = 2, 3, 4, 5, \dots$, focusing on the powers of n and r. First consider the case of $k = 2$. Here there is only one μ with $\mathrm{wt}(\mu) = 3$, namely $\mu = (2)$, which has $|\mu| = 2$ and $\ell(\mu) = 1$. Thus from $k = 2$ we pick up a factor on the order of $\frac{n^2}{r^3}$; more precisely, $\frac{n^{\downarrow 2}}{2r^3}$. This looks rather bad from the point of view of proving a quadratic lower bound for n: the term $\frac{n^{\downarrow 2}}{2r^3}$ is not small unless $n \ll r^{3/2}$. The main surprise in our proof is that this term will be exactly canceled by "lower-degree" contributions from larger k.

To see an example of this, consider the $k = 3$ contribution in (38). Here there are two μ's with $\mathrm{wt}(\mu) = 4$, namely $\mu = (3)$ and $\mu = (1, 1)$. The first gives a contribution on the order of $\frac{n^3}{r^5}$; more precisely, $-\frac{n^{\downarrow 3}}{3r^5}$. The second gives a contribution of $-\frac{n^{\downarrow 2}}{2r^3}$, thereby precisely canceling the $k = 2$ term. Thus we are left (so far) with $-\frac{n^{\downarrow 3}}{3r^5}$, which is small if $n \ll r^{5/3}$. This is still far from a quadratic bound, but it's better than the $r^{3/2}$ bound we were faced with previously.

In turn, the $-\frac{n^{\downarrow 3}}{3r^5}$ contribution will be canceled by a certain $k = 3$ term, namely $\frac{n^{\downarrow 3}}{r^5}$ from $\mu = (2, 1)$, together with a certain $k = 4$ term, namely $\frac{2n^{\downarrow 3}}{3r^5}$ from $\mu = (1, 1, 1)$. Indeed, if we sum up through $k = 6$, the total contribution is $-\frac{5n^{\downarrow 4}}{r^7} - \frac{n^{\downarrow 5}}{5r^9}$, which is small if $n \ll r^{7/4}$. This gets us still closer to a quadratic bound.

In fact, looking carefully at small partitions suggests that perfect cancelation is achieved if we group contributions according to $|\mu|$. This proves to be the case, as we will show below. In the end (38) does not precisely vanish because for $m > C/2$, not all μ's with $|\mu| = m$ appear in (38). However the "leftover contributions" are of the shape $r(\frac{n}{r^2})^k$ for $k > C/2$, a quantity we can ensure is small by taking ω and C large enough. (There is a tradeoff involved preventing us from taking C too large: our "error bound" (39) increases with C.)

5.2.4 Proof that the "main term" is small

Although (38) has a double summation, the summed quantity is simply counted exactly once for each μ with $3 \leq \mathrm{wt}(\mu) \leq C+1$. As suggested above, let us rearrange the summation according to $|\mu|$. We will use the notation $s = |\mu| - 1$ and $h = \ell(\mu) - 1$, so that $\mathrm{wt}(\mu) = s + h + 2$ (i.e., $k = s + h + 1$) and $\mathrm{wt}(\mu) \leq C + 1 \iff h \leq C - 1 - s$:

$$(38) = \sum_{s=1}^{C-1} \sum_{h=0}^{\min(s,C-1-s)} \sum_{\substack{\mu \vdash s+1 \\ \ell(\mu)=h+1}} \frac{(-1)^{s+h+1}}{(s+h+1)r^{s+h+1}} \frac{(s+h+1)^{\downarrow h}}{\mathrm{fact}(\mu)} n^{\downarrow(s+1)} r^{h-s}$$

$$= \sum_{s=1}^{C-1} (-1)^{s+1} \cdot \frac{n^{\downarrow(s+1)}}{r^{2s+1}} \sum_{h=0}^{\min(s,C-1-s)} (-1)^h (s+h)^{\downarrow(h-1)} \sum_{\substack{\mu \vdash s+1 \\ \ell(\mu)=h+1}} \frac{1}{\mathrm{fact}(\mu)}.$$

(We remark that we switched from $r + \frac{1}{2}$ to r in Lemma 5.3 so as to obtain nice cancelations on r here. We also recall the convention $m^{\downarrow(-1)} = \frac{1}{m+1}$.) It is not hard to show (see, e.g., [Mél10a, Lemma 11]) that

$$\sum_{\substack{\mu \vdash s+1 \\ \ell(\mu)=h+1}} \frac{1}{\mathrm{fact}(\mu)} = \frac{1}{(h+1)!} \binom{s}{h}.$$

Substituting this into the above, and also using $(s+h)^{\downarrow(h-1)} = \frac{(s+h)!}{(s+1)!}$, we get

$$(38) = \sum_{s=1}^{C-1} (-1)^{s+1} \cdot \frac{n^{\downarrow(s+1)}}{r^{2s+1}} \sum_{h=0}^{\min(s,C-1-s)} (-1)^h \frac{(s+h)!}{(s+1)!(h+1)!} \binom{s}{h}$$

$$= \sum_{s=1}^{C-1} \frac{(-1)^{s+1}}{s+1} \cdot \frac{n^{\downarrow(s+1)}}{r^{2s+1}} \sum_{h=0}^{\min(s,C-1-s)} \frac{(-1)^h}{h+1} \binom{s+h}{h} \binom{s}{h}.$$

We now obtain the promised cancelation. Specifically, it is a known combinatorial identity (see, e.g., [GKP94, page 182]) that for all $s \in \mathbb{Z}^+$, the inner summation equals 0 provided h ranges all the way up to s. In other words, all contributions from $s \leq \frac{C-1}{2}$ vanish. For larger s, it's not hard to bound the inner "partial sum" crudely by, say, 9^s in absolute value. We therefore finally conclude:

$$|(38)| \leq \sum_{\frac{C}{2} \leq s \leq C-1} \frac{1}{s+1} \cdot \frac{n^{\downarrow(s+1)}}{r^{2s+1}} \cdot 9^s \leq \frac{n}{r} \sum_{s \geq \frac{C}{2}} \left(\frac{9n}{r^2}\right)^s = \frac{r}{\omega^2} \sum_{s \geq \frac{C}{2}} \left(\frac{9}{\omega^2}\right)^s \leq r \left(\frac{3}{\omega}\right)^C. \tag{40}$$

5.2.5 Bounding the "error term"

In this section we bound the "error term" (39), using the following lemma:

Lemma 5.6. *Suppose* $n = \frac{r^2}{\omega^2}$. *Then* $0 \le \mathop{\mathbf{E}}_{\boldsymbol{\lambda} \sim \mathrm{SW}_r^n}[p_\mu^\sharp(\boldsymbol{\lambda})] \le r^{\mathrm{wt}(\mu)} \cdot (1/\omega^2)^{|\mu|}$.

Proof. By Corollary 2.35, $\mathop{\mathbf{E}}_{\boldsymbol{\lambda} \sim \mathrm{SW}_r^n}\left[p_\mu^\sharp(\boldsymbol{\lambda})\right] = n^{\downarrow|\mu|} r^{\ell(\mu)-|\mu|} \le n^{|\mu|} r^{\mathrm{wt}(\mu)-2|\mu|} = r^{\mathrm{wt}(\mu)} \cdot (1/\omega^2)^{|\mu|}$. □

We will first use this lemma to bound (39) in a "soft" way, thinking of C as an absolute universal constant. This is enough to get a testing lower bound like $n \ge \Omega_\delta(r^{2-\delta})$ for every $\delta > 0$. Subsequently we do some technical work (which the uninterested reader may skip) to get a more explicit lower bound.

Lemma 5.7. *For all* $C \ge 2$ *there is a constant* A_C *such that* $|(39)| \le A_C \cdot \frac{1}{\omega^2}$.

Proof. It suffices to show that for all $k \ge 2$ there is a constant A'_k such that

$$\frac{\mathbf{E}_{\boldsymbol{\lambda} \sim \mathrm{SW}_r^n}[\mathcal{O}_k(\boldsymbol{\lambda})]}{r^k} \le A'_k \cdot \frac{1}{\omega^2}.$$

But recalling Proposition 5.5, the left-hand side is

$$\sum_{\mu \,:\, \mathrm{wt}(\mu) \le k} c_{k,\mu} \mathop{\mathbf{E}}_{\boldsymbol{\lambda} \sim \mathrm{SW}_r^n}\left[\frac{p_\mu^\sharp(\boldsymbol{\lambda})}{r^k}\right],$$

and each expectation here is at most $(\frac{1}{\omega^2})^{|\mu|} \le \frac{1}{\omega^2}$ by Lemma 5.6. This completes the proof. □

Lemma 5.8. *In fact, the constants* A_C *from Lemma 5.7 satisfy* $A_C \le 2^{O(C^2 \log C)}$.

Proof. The proof involves some tedious analysis using the results of Section 2.8.1. It suffices to show that

$$\sum_{\mu : \mathrm{wt}(\mu) \le k} |c_{k,\mu}| \le 2^{O(k^2 \log k)}, \tag{41}$$

where, recall, the coefficients $c_{k,\mu}$ are defined by

$$p_k^* = \sum_{\mu \,:\, \mathrm{wt}(\mu) = k+1} \frac{k^{\downarrow(\ell(\mu)-1)}}{\mathrm{fact}(\mu)} p_\mu^\sharp + \sum_{\mu \,:\, \mathrm{wt}(\mu) \le k} c_{k,\mu} p_\mu^\sharp. \tag{42}$$

Let us return to the relationship between the p^* and $p^\sharp$ polynomials described in Section 2.8.1. Specifically, we'll need identities (7), (8), which express each $p_k^\sharp$ as a polynomial in $p_1^*, \dots, p_k^*$ via the power series $Q_k(t)$.

Given any polynomial R in indeterminates $p_1, \dots, p_k$ (either p^*'s or $p^\sharp$'s), write $\|R\|$ for the sum of the absolute values of R's coefficients. This is a submultiplicative norm. Observe from (8) that $\|Q_{k,m}\| \le (k+1)^{m+1}$ (indeed, one may show it's precisely $\frac{(k+1)^{m+1} - k^{m+1} - 1}{m+1}$). Thus the coefficient on t^s in $Q_k(t)^i$ is a polynomial in $p_1^*, \dots, p_k^*$ of norm at most $O(k)^s$. Hence the same is true for the coefficient on t^s in the expression $\sum_{i=0}^\infty \frac{(-1)^i}{i!} Q_k(t)^i$ from (8). As the coefficient on each power of t in $\prod_{j=1}^k (1 - (j - \frac{1}{2})t)$ is a number of magnitude at most $(k - \frac{1}{2})^k$, we finally deduce that the relationship (10) can be expressed more quantitatively as

$$p_k^\sharp = p_k^* + R_k(p_1^*, \dots, p_{k-1}^*), \quad \text{where } 1 + \|R_k\| \le \exp(bk \log k), \quad b \text{ a universal constant.}$$

We inductively invert this relationship as in (11), writing

$$p_k^* = S_k(p_1^\sharp, \dots, p_k^\sharp), \quad \text{where } S_k = p_k^\sharp + \Big\{\text{polynomial in } p_1^\sharp, \dots, p_{k-1}^\sharp \text{ of gradation at most } k-1\Big\}. \tag{43}$$

If we let $s(k) = \|S_k\|$, using convexity of $\exp(bk\log k)$ we get the inductive bound

$$s(k) \le \exp(bk\log k)s(k-1),$$

leading to the bound $s(k) \le \exp(O(k^2\log k))$. This is nearly enough to complete the proof; the only issue is that in (43) we have a polynomial in the $p_j^\sharp$'s, whereas in (42) we have the products of $p_j^\sharp$'s expanded out into linear combinations of $p_\mu^\sharp$'s. However Lemma 5.9 below, which crudely bounds the magnitude of the structure constants for the $p^\sharp$'s, shows that each monomial $\prod_i p_{\lambda_i}^\sharp$ with gradation $|\lambda| = w$ can be replaced by a linear polynomial in $p_\mu^\sharp$'s (with $|\mu| \le w$) wherein each coefficient has magnitude at most $4^{w^2\log w}$. Since w is always bounded by $k-1$ and since there are at most $2^{O(\sqrt{k})} \ll \exp(O(k^2\log k))$ partitions μ with $|\mu| \le k$, we conclude that each of these linear polynomials has norm at most $\exp(O(k^2\log k))$. Thus making these replacements in S_k only increases its norm by another multiplicative factor of $\exp(O(k^2\log k))$. The proof is complete. □

Lemma 5.9. *Let $\lambda \vdash w$, and suppose $\prod_{i=1}^{\ell(\lambda)} p_{\lambda_i}^\sharp = \sum_\mu c_\mu p_\mu^\sharp$ within Λ^*. Then $|c_\mu| \le 4^{w^2\log w}$ for all μ.*

Proof. The proof is an induction on $\ell = \ell(\lambda)$, the base case of $\ell = 1$ being trivial. Now for general λ with $\lambda_\ell = k$ we have

$$\prod_{i=1}^{\ell} p_{\lambda_i}^\sharp = \left(\prod_{i=1}^{\ell-1} p_{\lambda_i}^\sharp\right)\cdot p_k^\sharp = \left(\sum_\mu d_\mu p_\mu^\sharp\right)\cdot p_k^\sharp = \sum_\mu d_\mu \sum_\tau f_{\mu k}^\tau p_\tau^\sharp = \sum_\tau p_\tau^\sharp \sum_\mu d_\mu f_{\mu k}^\tau, \tag{44}$$

where each $|d_\mu|$ is at most $4^{(w-k)^2\log(w-k)} \le 4^{(w-1)^2\log(w)}$ by induction. By Corollary 2.37, the structure constants $f_{\mu k}^\tau$ satisfy $|f_{\mu k}^\tau| \le |C_{|\mu|k}^{|\tau|}| \le |\mu|!k! \le w^w$. Since the number of partitions of $(w-k)$ is trivially at most w^w, the coefficient on $p_\tau^\sharp$ in (44) has magnitude at most

$$\sum_\mu |d_\mu f_{\mu k}^\tau| \le w^{2w}\cdot \max_\mu |d_\mu| \le w^{2w}\cdot 4^{(w-1)^2\log(w)} \le 4^{w^2\log w},$$

completing the induction. □

5.2.6 Combining the bounds

Combining (40), and Lemmas 5.7, 5.8, we get that under the hypotheses of Proposition 5.4,

$$d_{\mathrm{KL}}(\mathrm{SW}_r^n, \mathrm{SW}_{r_+}^n) \le r\left(\frac{3}{\omega}\right)^C + \exp(O(C^2\log C))\cdot\frac{1}{\omega^2} + \tfrac{203}{\omega^3} \le \exp(O(C^{2.01}))\cdot\frac{1}{\omega^2}. \tag{45}$$

In the above we used $r\left(\frac{3}{\omega}\right)^C \le r\left(\frac{10}{\omega}\right)^C \le r\left(\frac{1}{10r}\right)^3 \le \frac{1}{\omega^3}$, the second inequality here following from the assumed lower bound on C. It's now evident that we should take C as small as we can; in particular, to equal $\lceil 3\frac{\log(10r)}{\log(\omega/10)}\rceil$. We conclude:

Theorem 5.10. *For any* $200 \le \omega \le \sqrt{r}$*, if* $n = \frac{r^2}{\omega^2}$ *then*

$$d_{\mathrm{KL}}(\mathrm{SW}_r^n, \mathrm{SW}_{r+1}^n) \le \exp(O((\log r)/(\log \omega))^{2.01}) \cdot \omega^{-2}.$$

In particular, for $\omega = \exp(O(\log^{.67} r))$ *and hence* $n = r^{2-O(1/\log^{.33} r)}$*, the above bound is* $o_r(1)$*.*

By Pinsker's inequality we may conclude also that $d_{\mathrm{TV}}(\mathrm{SW}_r^n, \mathrm{SW}_{r+}^n) \le o_r(1)$ unless $n = r^{2-O(1/\log^{.33} r)} = \widetilde{\Omega}(r^2)$. This completes the proof of the rank-r versus rank-$(r+1)$ testing lower bound; in particular, the more precise bound (27) in the case $\Delta = 1$.

5.3 Extension to $\Delta > 1$

Let us henceforth fix the parameter $C = \lceil 3\frac{\log(10r)}{\log(\omega/10)} \rceil$. To recap the preceding section we saw that

$$|\mathbf{E}[L_C(\boldsymbol{\lambda})]| \le \exp(O(C^{2.01})) \cdot \frac{1}{\omega^2}, \quad \text{and hence} \quad d_{\mathrm{KL}}(\mathrm{SW}_r^n, \mathrm{SW}_{r+1}^n) \le \exp(O(C^{2.01})) \cdot \frac{1}{\omega^2}. \tag{46}$$

If we apply Pinsker's inequality to the latter bound we obtain

$$d_{\mathrm{TV}}(\mathrm{SW}_r^n, \mathrm{SW}_{r+1}^n) \le \exp(O(C^{2.01})) \cdot \frac{1}{\omega}.$$

The key to getting a good lower bound when $\Delta > 1$ is to show that Pinsker's inequality is not sharp in our setting, and in fact the following is true:

Theorem 5.11. *For any* $200 \le \omega \le \sqrt{r}$*, if* $n = \frac{r^2}{\omega^2}$ *then*

$$d_{\mathrm{TV}}(\mathrm{SW}_r^n, \mathrm{SW}_{r+1}^n) \le \exp(O(C^{2.01})) \cdot \frac{1}{\omega^2}.$$

From this we can obtain the testing bound (27) for rank-r versus rank-$(r+\Delta)$ (where $1 \le \Delta \le r$) simply by using the triangle inequality. Specifically, given $r \le r' \le 2r$ and n, define $\omega_{r'}$ by $n = \frac{(r')^2}{\omega_{r'}^2}$. Applying Theorem 5.11 for each r', we get

$$d_{\mathrm{TV}}(\mathrm{SW}_{r'}^n, \mathrm{SW}_{r'+1}^n) \le \exp(O((\log r')/(\log \omega_{r'}))^{2.01}) \cdot \frac{1}{\omega_{r'}^2} \quad \text{for all } r \le r' < 2r.$$

But $\omega_{r'}$ is within a factor of 2 of ω_r for all $r \le r' \le 2r$; thus by adjusting the constant in the $O(\cdot)$, the above holds with ω_r in place of $\omega_{r'}$. Applying the triangle inequality, we get

$$d_{\mathrm{TV}}(\mathrm{SW}_r^n, \mathrm{SW}_{r'+\Delta}^n) \le \exp(O((\log r)/(\log \omega_r))^{2.01}) \cdot \frac{1}{\omega_r^2} \cdot \Delta.$$

Again, taking $\omega_r = \exp(O(\log^{.67} r))$, we get

$$d_{\mathrm{TV}}(\mathrm{SW}_r^n, \mathrm{SW}_{r'+\Delta}^n) \le \frac{n}{r^{2-O(1/\log^{.33} r)}} \cdot \Delta,$$

and this completes the proof of the rank-testing lower bound (27).

Thus it remains to prove Theorem 5.11. The main result we need for this is the following:

Theorem 5.12. $\operatorname*{\mathbf{Var}}_{\boldsymbol{\lambda} \sim \mathrm{SW}_r^n}[L_C(\boldsymbol{\lambda})] \le \exp(O(C^{2.01})) \cdot \frac{1}{\omega^4}.$

To prove Theorem 5.12 we will employ the following lemma:

Lemma 5.13. *Let μ be a partition with $\mathrm{wt}(\mu) = k \geq 2$. Then*

$$\operatorname*{\mathbf{Var}}_{\boldsymbol{\lambda} \sim \mathrm{SW}_r^n}[p_\mu^\sharp(\boldsymbol{\lambda})] \leq \exp(O(k^2 \log k)) \cdot r^{2k-2} \cdot (1/\omega^4).$$

Proof. If $|\mu| = 1$ then $p_\mu^\sharp(\lambda) = n$ which has variance 0. Thus we may assume $|\mu| \geq 2$ and hence $k \geq 3$. Using Proposition 2.39,

$$\mathbf{Var}[p_\mu^\sharp(\boldsymbol{\lambda})] = \mathbf{E}[p_\mu^\sharp(\boldsymbol{\lambda})^2] - \mathbf{E}[p_\mu^\sharp(\boldsymbol{\lambda})]^2 = \mathbf{E}[p_{\mu \cup \mu}^\sharp(\boldsymbol{\lambda})] - \mathbf{E}[p_\mu^\sharp(\boldsymbol{\lambda})]^2 + \mathbf{E}[q_\mu(\boldsymbol{\lambda})] \tag{47}$$

where $q_\mu(\lambda)$ is a certain linear combination of $p_\nu^\sharp$ polynomials, each of weight at most $2k-2$. Regarding the first two quantities here, Corollary 2.35 tells us that

$$\mathbf{E}[p_{\mu\cup\mu}^\sharp(\boldsymbol{\lambda})] - \mathbf{E}[p_\mu^\sharp(\boldsymbol{\lambda})]^2 = n^{\downarrow(2|\mu|)} r^{2\ell(\mu)-2|\mu|} - (n^{\downarrow|\mu|} r^{\ell(\mu)-|\mu|})^2 = r^{2\ell(\mu)-2|\mu|}(n^{\downarrow(2|\mu|)} - (n^{\downarrow|\mu|})^2),$$

which is evidently nonpositive. Thus it suffices to prove the upper bound

$$|\mathbf{E}[q_\mu(\boldsymbol{\lambda})]| \leq \exp(O(k^2 \log k)) \cdot r^{2k-2} \cdot (1/\omega^4). \tag{48}$$

By Lemma 5.9, the coefficients on the $p_\nu^\sharp$'s in the linear combination $q_\mu(\lambda)$ each have magnitude at most $\exp(O(k^2 \log k))$, and there are at most $2^{O(\sqrt{k})}$ of them. Thus (48) follows provided we can show $\mathbf{E}[p_\nu^\sharp(\boldsymbol{\lambda})] \leq r^{2k-2}/\omega^4$ for all ν of weight at most $2k-2$. This is immediate from Lemma 5.6 for all $\nu \neq (1)$, and when $\nu = (1)$ it still holds: Lemma 5.6 gives us the bound $r^2/\omega^2 \leq r^3/\omega^4 \leq r^{2k-2}/\omega^4$, the first inequality using $\omega \leq \sqrt{r}$ and the second using $k \geq 3$. □

We can now prove Theorem 5.12.

Proof of Theorem 5.12. Recall identity (37):

$$L_C(\lambda) = \sum_{k=2}^{C} \frac{(-1)^k}{kr^k} \cdot \sum_{\mathrm{wt}(\mu)=k+1} \frac{k^{\downarrow(\ell(\mu)-1)}}{\mathrm{fact}(\mu)} p_\mu^\sharp(\lambda) + \sum_{k=2}^{C} \frac{(-1)^k \mathcal{O}_k(\lambda)}{kr^k}.$$

We claim that for each $2 \leq k \leq C$,

$$\mathbf{Var}\left[\frac{(-1)^k \mathcal{O}_k(\boldsymbol{\lambda})}{kr^k}\right] \leq \exp(O(C^{2.01})) \cdot \frac{1}{\omega^4}, \tag{49}$$

and that furthermore for each μ of weight $k+1$ we have

$$\mathbf{Var}\left[\frac{(-1)^k}{kr^k} \cdot \frac{k^{\downarrow(\ell(\mu)-1)}}{\mathrm{fact}(\mu)} p_\mu^\sharp(\boldsymbol{\lambda})\right] \leq \exp(O(C^{2.01})) \cdot \frac{1}{\omega^4}. \tag{50}$$

This is sufficient to complete the proof, as in general

$$\mathbf{Var}[\boldsymbol{X}_1 + \cdots + \boldsymbol{X}_m] \leq m(\mathbf{Var}[\boldsymbol{X}_1] + \cdots + \mathbf{Var}[\boldsymbol{X}_m]); \tag{51}$$

in our particular case we have only $m = \exp(O(\sqrt{C}))$ summands, and this factor can be absorbed into the target variance bound of $\exp(O(C^{2.01})) \cdot (1/\omega^4)$. To verify (49), first recall that each $\mathcal{O}_k(\boldsymbol{\lambda})$ is a linear combination of $p_\nu^\sharp(\lambda)$'s for ν of weight at most $k \leq C$; further, the sum of the absolute

value of the coefficients is at most $\exp(O(C^{2.01}))$ (see (41)). Using (51) again, it therefore suffices to check that

$$\mathbf{Var}\left[\frac{p^\sharp_\nu(\boldsymbol{\lambda})}{r^k}\right] \le \exp(O(C^{2.01})) \cdot \frac{1}{\omega^4}$$

when $\mathrm{wt}(\nu) \le k \le C$. By Lemma 5.13 this is true, with a factor of r^{-2} to spare.

To verify (50), we may ignore the factor $\frac{(-1)^k}{k \cdot \mathrm{fact}(\mu)}$, and also ignore the factor $k^{\downarrow(\ell(\mu)-1)}$ as it contributes at most a multiplicative $C^C \ll \exp(O(C^{2.01}))$. Thus it suffices to show $\mathbf{Var}[p^\sharp_\mu(\boldsymbol{\lambda})/r^k] \le \exp(O(C^{2.01}))/\omega^4$ for μ of weight $k+1$ (and $k \le C$). But this is immediate from Lemma 5.13. □

We now work towards the proof of Theorem 5.11. Adding Theorem 5.12 and the square of (46) we obtain

$$\mathop{\mathbf{E}}_{\boldsymbol{\lambda} \sim \mathrm{SW}^n_r}[L_C(\boldsymbol{\lambda})^2] \le \exp(O(C^{2.01})) \cdot \frac{1}{\omega^4}. \tag{52}$$

We would now like to similarly claim that

$$\mathop{\mathbf{E}}_{\boldsymbol{\lambda} \sim \mathrm{SW}^n_{r_+}}[L^+_C(\boldsymbol{\lambda})^2] \le \exp(O(C^{2.01})) \cdot \frac{1}{\omega^4}, \tag{53}$$

where we are writing

$$L^+_C(\lambda) \coloneqq \sum_{k=2}^{C} \frac{(-1)^k p^*_k(\lambda)}{k(r+1)^k}.$$

To obtain this, it suffices to repeat all of the arguments beginning with Section 5.2.2 until this point; the only thing that really changes is that $\omega = \omega_r$ needs to be replaced with ω_{r+1}, but this has a negligible effect on the bounds (and indeed usually very slightly improves them).

Next, we claim that Lemma 5.3 continues to hold if we replace $L_C(\lambda)$ with the analogous $L^+_C(\lambda)$. The key change to the proof comes in the last main inequality, where we need to observe that the

$$\left(\frac{1}{r^k} - \frac{1}{(r+\frac{1}{2})^k}\right) \le \frac{k}{2r^{k+1}}$$

continues to hold if the left-hand side is replaced with

$$\left(\frac{1}{(r+\frac{1}{2})^k} - \frac{1}{(r+1)^k}\right).$$

We need one more definition for the proof of Theorem 5.11.

Definition 5.14. Say that $\lambda \vdash n$ is *usual*$^+$ if it is usual and if furthermore $|L^*_\infty(\lambda)| \le 2$.

Lemma 5.15. *Both for $\boldsymbol{\lambda} \sim \mathrm{SW}^n_r$ and $\boldsymbol{\lambda} \sim \mathrm{SW}^n_{r_+}$ it holds that*

$$\mathbf{Pr}[\boldsymbol{\lambda} \text{ not usual}^+] \le \exp(O(C^{2.01})) \cdot \frac{1}{\omega^4}.$$

Proof. For $\boldsymbol{\lambda} \sim \mathrm{SW}^n_r$, Lemma 5.2 tells us that

$$\mathbf{Pr}[\boldsymbol{\lambda} \text{ not usual}] \le 2^{-20r/\omega} \le 2^{-\Omega(\sqrt{r})} \ll \exp(O(C^{2.01})) \cdot \frac{1}{\omega^4}$$

and it's easy to check that this is also true with plenty of room to spare for $\boldsymbol{\lambda} \sim \mathrm{SW}^n_{r_+}$. Thus it suffices to verify for both distributions on $\boldsymbol{\lambda}$ that the probability of $|L^*_\infty(\lambda)| \le 2$ satisfies the same upper bound. By applying Markov's inequality to (52), (53) we get

$$\mathop{\mathbf{Pr}}_{\boldsymbol{\lambda} \sim \mathrm{SW}^n_r}[L_C(\boldsymbol{\lambda})^2 \ge 1], \mathop{\mathbf{Pr}}_{\boldsymbol{\lambda} \sim \mathrm{SW}^n_{r_+}}[L^+_C(\boldsymbol{\lambda})^2 \ge 1] \le \exp(O(C^{2.01})) \cdot \frac{1}{\omega^4}.$$

Finally, when $\boldsymbol{\lambda}$ is usual and $|L_C(\boldsymbol{\lambda})^2| \not\ge 1$, it follows that necessarily $|L^*_\infty(\boldsymbol{\lambda})| \le 2$, in light of Lemma 5.3 and the fact that $\frac{201}{\omega^3} \le 1$. As noted earlier, the r_+-analogue of Lemma 5.3 holds, and hence we may draw the same conclusion concerning $L^+_C(\boldsymbol{\lambda})^2$. □

Finally we are ready to complete the proof of Theorem 5.11. We begin with

$$d_{\mathrm{TV}}(\mathrm{SW}^n_r, \mathrm{SW}^n_{r_+}) \le \frac{1}{2} \mathop{\mathbf{Pr}}_{\boldsymbol{\lambda} \sim \mathrm{SW}^n_r}[\boldsymbol{\lambda} \text{ not usual}^+] + \frac{1}{2} \mathop{\mathbf{Pr}}_{\boldsymbol{\lambda} \sim \mathrm{SW}^n_{r_+}}[\boldsymbol{\lambda} \text{ not usual}^+] + \frac{1}{2} \sum_{\text{usual}^+ \ \lambda} \left| \mathrm{SW}^n_{r_+}[\lambda] - \mathrm{SW}^n_r[\lambda] \right|.$$

We can bound the first two terms above using Lemma 5.15. Indeed there is room to spare, as the bound we get is the square of what we can tolerate. Thus it remains to bound the third term by $\exp(O(C^{2.01})) \cdot \frac{1}{\omega^2}$. For it we use

$$\begin{aligned} \sum_{\text{usual}^+ \ \lambda} \left| \mathrm{SW}^n_{r_+}[\lambda] - \mathrm{SW}^n_r[\lambda] \right| &= \mathop{\mathbf{E}}_{\boldsymbol{\lambda} \sim \mathrm{SW}^n_{r_+}} \left[1_{\{\boldsymbol{\lambda} \text{ usual}^+\}} \cdot \left| 1 - \frac{\mathrm{SW}^n_r[\boldsymbol{\lambda}]}{\mathrm{SW}^n_{r_+}[\boldsymbol{\lambda}]} \right| \right] \\ &= \mathop{\mathbf{E}}_{\boldsymbol{\lambda} \sim \mathrm{SW}^n_{r_+}} \left[1_{\{\boldsymbol{\lambda} \text{ usual}^+\}} \cdot |1 - \exp(u(\boldsymbol{\lambda}))| \right] \end{aligned} \tag{54}$$

where

$$u(\boldsymbol{\lambda}) = \ln \left(\frac{\mathrm{SW}^n_r[\lambda]}{\mathrm{SW}^n_{r_+}[\lambda]} \right) = n \ln \left(1 + \frac{1}{r} \right) - \frac{n}{r + \frac{1}{2}} + L^*_\infty(\boldsymbol{\lambda}), \tag{55}$$

the last equality holding from (31) (see also the sentence after (33)) under the assumption that $\boldsymbol{\lambda}$ is usual (which we can indeed assume, since we're multiplying against $1_{\{\boldsymbol{\lambda} \text{ usual}^+\}}$). As we noted after (33), the first two quantities in (55) sum to a positive quantity not exceeding $\frac{n}{12r^3} \le \frac{1}{\omega^2}$. Furthermore, because of the presence of the usual$^+$-indicator in (54) we may assume in analyzing (55) that $|L^*_\infty(\boldsymbol{\lambda})| \le 2$. Thus we may use the bound $u(\boldsymbol{\lambda}) \le 2 + \frac{1}{\omega^2} \le 2.01$. Since $|1 - \exp(u)| \le 4|u|$ for $u \in [-2.01, 2.01]$, we may conclude that

$$(54) \le 4 \mathop{\mathbf{E}}_{\boldsymbol{\lambda} \sim \mathrm{SW}^n_{r_+}} \left[1_{\{\boldsymbol{\lambda} \text{ usual}^+\}} \cdot \left(\frac{1}{\omega^2} + |L^*_\infty(\boldsymbol{\lambda})| \right) \right].$$

Thus to complete the proof of Theorem 5.11 it remains to show

$$\mathop{\mathbf{E}}_{\boldsymbol{\lambda} \sim \mathrm{SW}^n_{r_+}} [|L^*_\infty(\boldsymbol{\lambda})|] \le \exp(O(C^{2.01})) \cdot \frac{1}{\omega^2}.$$

By the r_+-analogue of Lemma 5.3, it suffices to prove this with $L^+_C(\boldsymbol{\lambda})$ in place of $L^*_\infty(\boldsymbol{\lambda})$, because $201/\omega^3 \ll \exp(O(C^{2.01}))/\omega^2$. But finally

$$\mathop{\mathbf{E}}_{\boldsymbol{\lambda} \sim \mathrm{SW}^n_{r_+}} \left[|L^+_C(\boldsymbol{\lambda})| \right] \le \sqrt{\mathop{\mathbf{E}}_{\boldsymbol{\lambda} \sim \mathrm{SW}^n_{r_+}} \left[L^+_C(\boldsymbol{\lambda})^2 \right]} \le \exp(O(C^{2.01})) \cdot \frac{1}{\omega^2},$$

using Cauchy–Schwarz and (53). The proof of Theorem 5.11—and hence also the testing lower bound (27)—is therefore complete.

6 Quantum rank testing

6.1 Testers with one-sided error

In this section, we prove the first part of Theorem 1.11, that $\Theta(r^2/\epsilon)$ copies are necessary and sufficient to test whether or not a state has rank r with one-sided error. We will show this by analyzing the following algorithm.

Rank Tester. Given $\rho^{\otimes n}$,

1. Sample $\boldsymbol{\lambda} \sim \mathrm{SW}_\rho^n$.
2. Accept if $\ell(\boldsymbol{\lambda}) \leq r$. Reject otherwise.

Our primary tool in analyzing this tester will be the RSK correspondence. Suppose ρ's nonzero eigenvalues are $\eta = \{\eta_1, \ldots, \eta_d\}$, and let $\mathcal{D}$ be the distribution over $[d]$ induced by η. By Remark 2.24, SW_ρ^n has the same distribution as the process which first samples $\boldsymbol{w} \sim \mathcal{D}^{\otimes n}$ and outputs $\boldsymbol{\lambda} = \mathrm{RSK}(\boldsymbol{w})$. Write $\mathrm{LDS}(\boldsymbol{w})$ for the length of the longest strongly decreasing subsequence in $\boldsymbol{w}$. By Theorem 2.14, $\ell(\boldsymbol{\lambda}) = \mathrm{LDS}(\boldsymbol{w})$.

The key property we will need of the Rank Tester is the following:

Proposition 6.1. *The Rank Tester is the optimal algorithm for testing whether or not a state has rank r with one-sided error.*

Proof. To show this, we need to show (i) that every λ satisfying $\ell(\lambda) \leq r$ occurs with nonzero probability in SW_ρ^n for some ρ of rank r and (ii) that no λ satisfying $\ell(\lambda) > r$ occurs in SW_ρ^n for any ρ of rank r. The first follows because if ρ has r nonzero eigenvalues, then the word

$$w \coloneqq \underbrace{r, \ldots, r}_{\lambda_r \text{ letters}}, \underbrace{(r-1), \ldots, (r-1)}_{\lambda_{r-1} \text{ letters}}, \ldots, \underbrace{1, \ldots, 1}_{\lambda_1 \text{ letters}}$$

occurs in $\mathcal{D}^{\otimes n}$ with nonzero probability. It is easy to check that $\lambda = \mathrm{RSK}(w)$.

To show that (ii) holds, if ρ is rank r, then η has at most r nonzero entries. Thus, any word w in the support of $\mathcal{D}^{\otimes n}$ will always satisfy $\mathrm{LDS}(w) \leq r$ because w will contain at most r distinct letters. As $\ell(\lambda) = \mathrm{LDS}(w)$, we are done. □

As a result of Proposition 6.1, Theorem 1.11 follows from the following lemma.

Lemma 6.2. *The Rank Tester tests whether or not a state has rank r with $\Theta(r^2/\epsilon)$ copies.*

Proof. If ρ is ϵ-far from having rank r, then η is ϵ-far in TV distance from having support size r. Thus, we can show the lemma by showing the following two facts about probability distributions.

(i) For every probability distribution $\mathcal{D} = (p_1, \ldots, p_d)$ which is ϵ-far from having support size r, a random word $\boldsymbol{w} \sim \mathcal{D}^{\otimes n}$ satisfies $\mathrm{LDS}(\boldsymbol{w}) \geq r + 1$ with probability at least $2/3$ for some $n = O(r^2/\epsilon)$.

(ii) There exists an integer d and a probability distribution $\mathcal{D} = (p_1, \ldots, p_d)$ which is ϵ-far from having support size r such that, for a random word $\boldsymbol{w} \sim \mathcal{D}^{\otimes n}$, $\mathrm{LDS}(\boldsymbol{w}) \leq r$ with probability greater than $1/3$ whenever $n = o(r^2/\epsilon)$.

Proof of statement (i): We will need the following concentration bound for sums of geometric random variables.

Proposition 6.3 ([Bro]). *Write $X = X_1 + \ldots + X_n$, where the X_i's are iid geometric random variables with expectation μ. For any $k > 1$, $\mathbf{Pr}[X > kn\mu] \leq \exp\left(-\frac{1}{2}kn(1-1/k)^2\right)$.*

We note that Proposition 6.3 also holds with the weaker hypothesis that the X_i's are independent (and not necessarily identically distributed), each with expectation at most μ.

We may assume that $p_1 \geq \ldots \geq p_d$. We will split into two cases, handled below: (1) $p_{r+1} \geq \epsilon/4r$ and (2) $p_{r+1} < \epsilon/4r$.

(1) Because the probabilities are sorted, $p_1, \ldots, p_{r+1} \geq \epsilon/4r$. For the infinite random word $\boldsymbol{w} \sim \mathcal{D}^{\otimes\infty}$, consider the number of letters one has to traverse through before finding $(r+1), r, \ldots, 1$ as a subsequence. This number is distributed as $\boldsymbol{X} = \boldsymbol{X}_{r+1} + \ldots + \boldsymbol{X}_1$, where $\boldsymbol{X}_i$ is a geometric random variable with success probability p_i.

By assumption, $p_i \geq \epsilon/4r$, and therefore $\mathbf{E}\,\boldsymbol{X}_i \leq 4r/\epsilon$, for each $i \in [r+1]$. By Proposition 6.3, $\boldsymbol{X}$ is at most $24r^2/\epsilon$ with probability at least $2/3$. Thus, if $n = 24r^2/\epsilon$, then $\boldsymbol{w} \sim \mathcal{D}^{\otimes n}$ has a strictly decreasing subsequence of size $r+1$ with high probability.

(2) Because the probabilities are sorted, $p_{r+1}, \ldots, p_d < \epsilon/4r$. Place the letters from $\{r+1, \ldots, d\}$ into buckets as follows: starting from letter $(r+1)$ and proceeding in order, add each letter to the current bucket until it contains at least $\epsilon/4r$ weight. At this point, move to the next bucket and repeat this process starting with the current letter until all letters have been bucketed.

Because these letters have weight $\leq \epsilon/4r$, each bucket has total weight in the interval $[\epsilon/4r, \epsilon/2r)$ (except possibly the final bucket). There must be at least $2r+1$ buckets with nonzero total weight, as otherwise $p_{r+1} + \ldots + p_d < \epsilon$, contradicting the fact that $\mathcal{P}$ is ϵ-far from having support size r. This gives us at least $2r \geq r+1$ buckets each of which contains at least $\epsilon/4r$ total weight.

Now we can use an argument similar to case (1) to show that when $n = 24r^2/\epsilon$, a random $\boldsymbol{w} \sim \mathcal{D}^{\otimes n}$ will with probability $\geq 2/3$ have a strictly decreasing subsequence in which the first letter comes from bucket $r+1$, the second letter comes from bucket r, and so on (ending in a letter from the first bucket). This is a strictly decreasing subsequence of size $r+1$.

Proof of statement (ii): For $d \gg r$, define the probability distribution

$$\mathcal{P} = \left(1 - 2\epsilon, \frac{2\epsilon}{d-1}, \ldots, \frac{2\epsilon}{d-1}\right).$$

Because $d \gg r$, $\mathcal{P}$ is ϵ-far from having support size r. For a string $w \in [d]^n$, let $\widetilde{w}$ be the substring of w formed by deleting all occurrences of the letter "1" from w. It is easy to see that $\mathrm{LDS}(\widetilde{w}) \leq \mathrm{LDS}(w) \leq \mathrm{LDS}(\widetilde{w}) + 1$.

For a randomly drawn $\boldsymbol{w} \sim \mathcal{P}^{\otimes n}$, let us condition on $\widetilde{\boldsymbol{w}}$ having a certain fixed length m. The value of $\mathrm{LDS}(\widetilde{\boldsymbol{w}})$ is distributed as the length of the longest decreasing subsequence in a uniformly random word drawn from $[d-1]^m$. By Theorem 2.14, this is distributed as $\boldsymbol{\lambda}'_1$ for $\boldsymbol{\lambda} \sim \mathrm{SW}^m_{d-1}$. Setting $B = \lceil 100\sqrt{m} \rceil$, let us show that $\mathbf{Pr}[\boldsymbol{\lambda}'_1 \geq B]$ is small. If $B \geq d$, then surely $\boldsymbol{\lambda}'_1 < B$ always, as $\boldsymbol{\lambda} \sim \mathrm{SW}^m_{d-1}$ will always have height at most $d-1$. On the other hand, if $B < d$, then by Proposition 2.31,

$$\mathbf{Pr}[\boldsymbol{\lambda}'_1 \geq B] \leq \left(\frac{2e^2 m}{B^2}\right)^B \leq \frac{2e^2}{10000}.$$

In summary, conditioned on $\widetilde{\boldsymbol{w}}$ having a certain fixed length m, $\mathrm{LDS}(\widetilde{\boldsymbol{w}}) \leq O(\sqrt{m})$ with all but the above probability.

In expectation, for a random $\boldsymbol{w} \sim \mathcal{P}^{\otimes n}$, $\widetilde{\boldsymbol{w}}$ has length $2\epsilon d$. By Markov's inequality, the probability that the length of $\widetilde{\boldsymbol{w}}$ is greater than $200\epsilon d$ is at most $1/100$. Conditioned on the length of $\widetilde{\boldsymbol{w}}$ being at most $200\epsilon d$, the above paragraph tells us that $\mathrm{LDS}(\widetilde{\boldsymbol{w}}) \leq O(\sqrt{\epsilon d})$ with probability $1 - 2e^2/10000$. Thus, when $\boldsymbol{w} \sim \mathcal{P}^{\otimes n}$, we have with probability greater than $1/3$ that $\mathrm{LDS}(\boldsymbol{w}) \leq O(\sqrt{\epsilon d})$, which is $o(r)$ unless $d = \Omega(r^2/\epsilon)$. □

For our last result of this section, we will show that the copy complexity of the Rank Tester can be improved in certain interesting cases. In particular, the Rank Tester matches the upper bound of the Uniform Distribution Distinguisher from Section 5 for the case of r v. $r+1$, and does so with one-sided error.

Proposition 6.4. *The Rank Tester can distinguish between the case when ρ's spectrum is uniform on either r or $r+1$ eigenvalues with $O(r^2)$ copies of ρ.*

Proof. If ρ's spectrum is uniform on r eigenvalues, then it is rank r and so the Rank Tester never rejects. Thus, we need only show that the Rank Tester rejects with probability $\geq 2/3$ when ρ's spectrum is uniform on $r+1$ eigenvalues for some $n = O(r^2)$. We will follow the analysis in the proof of statement (i) above and show that a random word $\boldsymbol{w} \sim \mathcal{D}^{\otimes n}$ has $\mathrm{LDS}(\boldsymbol{w}) = r+1$ with high probability. The gain will come from the fact that $\eta = (1/(r+1), \ldots, 1/(r+1))$.

For the infinite random word $\boldsymbol{w} \sim \mathcal{D}^{\otimes \infty}$, consider the number of letters one has to traverse through before finding $(r+1), r, \ldots, 1$ as a subsequence. This number is distributed as $\boldsymbol{X} = \boldsymbol{X}_{r+1} + \ldots + \boldsymbol{X}_1$, where $\boldsymbol{X}_i$ is a geometric random variable with success probability $1/(r+1)$ and expectation $r+1$. By Proposition 6.3, $\boldsymbol{X}$ is at most $6r^2$ with probability at least $2/3$. Thus, if $n = 6r^2$, then $\boldsymbol{w} \sim \mathcal{D}^{\otimes n}$ has a strictly decreasing subsequence of size $r+1$ with high probability. □

6.2 A lower bound for testers with two-sided error

In this section, we prove the second part of Theorem 1.11, that $\Omega(r/\epsilon)$ copies are necessary to test whether or not a state has rank r with two-sided error.

Proof. Let $d \gg r$. In this proof, we will take the viewpoint of a density matrix as a probability distribution over pure states. Let ρ and σ be maximally mixed on subspaces of dimension $(r-1)$ and $(d-1)$, respectively. Consider the following process for generating a product state $|\Psi\rangle = |\Psi_1\rangle \otimes \cdots \otimes |\Psi_n\rangle$:

1. Let $x \in \{0,1\}^n_{2\epsilon}$ be a uniformly random 2ϵ-biased string, meaning each coordinate is selected independently according to $\mathbf{Pr}[x_i = 1] = 2\epsilon$.
2. For each $i \in [n]$ such that $x_i = 0$, set $|\Psi_1\rangle \coloneqq |d\rangle$.
3. Let b be an arbitrary $\{0,1\}$-bit. For each $i \in [n]$ such that $x_i = 1$,
   (a) if $b = 0$, then set $|\Psi_i\rangle$ to be a state vector sampled from ρ.
   (b) if $b = 1$, then set $|\Psi_i\rangle$ to be a state vector sampled from σ.

If b is 0, then the mixed state output by this procedure has spectrum $(1-2\epsilon, \frac{2\epsilon}{r-1}, \ldots, \frac{2\epsilon}{r-1})$, which is rank r. On the other hand, if b is 1, then the mixed state output by this procedure has spectrum $(1-2\epsilon, \frac{2\epsilon}{d-1}, \ldots, \frac{2\epsilon}{d-1})$, which because $d \gg r$ is ϵ-far from having rank r.

Let us consider the choice of x in the first step, and set $\mathrm{wt}(\boldsymbol{x})$ to be the number of 1's in x. In expectation, $\mathrm{wt}(\boldsymbol{x})$ will be $2\epsilon n$, and so by Markov's inequality $\mathrm{wt}(\boldsymbol{x})$ will be at most $200\epsilon n$ with probability at least $99/100$. There must exist an x with $\mathrm{wt}(x) \leq 200\epsilon n$ conditioned on which the algorithm succeeds with probability at least $3/5$, as otherwise it will succeed in total with probability at most $1/100 + 99/100 \cdot 3/5 < 2/3$.

Fix any such x. The job of the algorithm is reduced to distinguishing between the cases when those $|\Psi_i\rangle$'s for which $x_i = 1$ came from ρ which is maximally mixed on a subspace of dimension $(r-1)$ (when $b = 0$) or from σ which is maximally mixed on a subspace of dimension $(d-1)$ (when $b = 1$). Because $d \gg r$, we have by Theorem 1.9 that this requires at least $\Omega(r)$ copies to succeed with probability at least $3/5$. Thus, we must have $200\epsilon n \geq \Omega(r)$, in which case $n = \Omega(r/\epsilon)$. □

7 The EYD lower bound (continued)

In this section, we prove Theorem 3.4.

Theorem 3.4 restated. *For every constant $C > 0$, there are constants $\delta, \epsilon > 0$ such that*

$$\Pr_{\boldsymbol{\lambda} \sim \mathrm{SW}_d^n}[d_{\mathrm{TV}}(\underline{\boldsymbol{\lambda}}, \mathsf{Unif}_d) > \epsilon] \geq \delta$$

when $n \leq Cd^2$ and d is sufficiently large.

Proof. To prove Theorem 3.4, we show, at a high level, that when $n \leq Cd^2$, Biane's law of large numbers kicks in and $\overline{\boldsymbol{\lambda}}$ approaches the limiting curve Ω_θ, for $\theta := \frac{\sqrt{n}}{d}$. Each of these curves is constantly far from the curve produced by the uniform partition, and the lower bound follows. However, carrying out this proof involves some subtle argumentation and splitting of hairs which we will go into.

There is one regime where $\overline{\boldsymbol{\lambda}}$ certainly does not approach Ω_θ: when n is a fixed value independent of the value of d, then $\overline{\boldsymbol{\lambda}}$ will be always be constantly far from Ω_θ. However, we can rule this case out by noting that when n is too small as a function of d, then any $\lambda = (\lambda_1, \ldots, \lambda_d)$ with n boxes will have most of its λ_i's zero, and so $\underline{\lambda}$ will be far from uniform. In particular, when $n = o(d)$, then we have that $d_{\mathrm{TV}}(\underline{\boldsymbol{\lambda}}, \mathsf{Unif}_d) \to 1$ as $d \to \infty$. As a result, for sufficiently large d we can immediately assume that $n \geq f(d)$, where $f(d)$ is any function which is both $\omega_d(1)$ and $o(d)$. For concreteness, we will take $f(d) := \sqrt{d}$.

We are now in the regime where Biane's law of large numbers holds. Theorem 2.30 tells us that if $\frac{\sqrt{n}}{d} \sim c$ for c some absolute constant, then there is some constant $d(c) > 0$ such that for a random $\boldsymbol{\lambda} \sim \mathrm{SW}_d^n$, $\overline{\boldsymbol{\lambda}}$ is ϵ-close (in L^∞ distance) to Ω_c whenever $d \geq d(c)$. The main difficulty we have in applying Biane's law of large numbers directly is that the function $d(c)$ is left unspecified and, for example, could be wildly different even for two close values of c. This is problematic in our case, because for each value of d, the ratio $\theta = \frac{\sqrt{n}}{d}$ may be any real number in the interval $[\sqrt{f(d)}/d, \sqrt{C}]$, and so θ may jump around and never converge to a fixed value c. In particular, an adversary could potentially choose n (and therefore θ) as a function of d cleverly so that for each d, we have that $d < d(\theta)$, and so Biane's law of large numbers never applies. Though seemingly unlikely, this possibility is not ruled out by the statements of known theorems.

Our goal now is to show that the convergence to the limiting shapes guaranteed by Biane's theorem happens at roughly the same rate for all values of θ in our interval. First we will need a definition.

Definition 7.1. Given continual diagrams $f, g : \mathbb{R} \to \mathbb{R}$, the L^1 *distance* between them is

$$d_1(f, g) := \int_{\mathbb{R}} |f(x) - g(x)| \, \mathrm{d}x.$$

This defines a metric on the set of continual diagrams, and it is well-defined because $f(x) - g(x) = 0$ whenever $|x|$ is sufficiently large. If λ, μ are both partitions of n, then $d_1(\overline{\lambda}, \overline{\mu}) = 4 \cdot d_{\mathrm{TV}}(\underline{\lambda}, \underline{\mu},)$.

We will prove the following result:

Theorem 7.2. *Let $C > 0$ be an absolute constant, and let $f(d) : \mathbb{N} \to \mathbb{N}$ be $\omega_d(1)$. Then for any constant $0 < \delta < 1$, if $f(d) \le n \le Cd^2$, then*

$$\mathop{\mathbf{Pr}}_{\boldsymbol{\lambda} \sim \mathrm{SW}_d^n} \left[d_1(\overline{\boldsymbol{\lambda}}, \Omega_\theta) \ge \delta \right] \le \delta,$$

for sufficiently large d, where $\theta = \frac{\sqrt{n}}{d}$.

Let us now complete the argument assuming Theorem 7.2. For $\kappa > 0$, define the following continual diagram:

$$\overline{\mathsf{unif}_\kappa}(x) := \begin{cases} x + \frac{2}{\kappa} & \text{if } x \in (-\frac{1}{\kappa}, \kappa - \frac{1}{\kappa}] \\ -x + 2\kappa & \text{if } x \in (\kappa - \frac{1}{\kappa}, \kappa), \\ |x| & \text{otherwise.} \end{cases} \tag{56}$$

To see how such a function arises, consider the uniform "partition" $\left(\frac{n}{d}, \ldots, \frac{n}{d}\right)$ ("partition" being in quotation marks because $\frac{n}{d}$ may not be integral). Drawing this in the French notation gives a rectangle of width $\frac{n}{d}$ and height d whose bottom-left corner is the origin. Drawing this in the Russian notation and dilating by a factor of $1/\sqrt{n}$ therefore gives the curve $\overline{\mathsf{unif}_\theta}(x)$. One consequence of this is that if λ is a partition of n, then $d_1(\overline{\lambda}, \overline{\mathsf{unif}_\theta}) = 4 \cdot d_{\mathrm{TV}}(\underline{\lambda}, \mathsf{Unif}_d,)$.

Define the function $\Delta : (0, \sqrt{C}] \to \mathbb{R}^{\ge 0}$ by $\Delta(\kappa) := d_1(\overline{\mathsf{unif}_\kappa}, \Omega_\kappa)$. When $\kappa < .3$, $\Delta(\kappa) > .5$ for all c. This is because $\Omega_\kappa(x) = -x$ for all $x \le -2$ regardless of κ, whereas $\overline{\mathsf{unif}_\kappa}(x) = -x + 2\kappa$ in $(\kappa - \frac{1}{\kappa}, -2]$. Because $\kappa < .3$,

$$d_1(\overline{\mathsf{unif}_\kappa}, \Omega_\kappa) = \int_{\mathbb{R}} \left| \overline{\mathsf{unif}_\kappa}(x) - \Omega_\kappa(x) \right| \mathrm{d}x \ge 2\kappa \cdot \left(\tfrac{1}{\kappa} - 2 - \kappa \right) \ge 0.5.$$

Now, let us lower-bound $\Delta(\kappa)$ when $\kappa \ge .3$. Write I for the interval $[.3, \sqrt{C}]$. (If $.3 > \sqrt{C}$ then this step can be skipped.) To begin, we note that $\Delta(\kappa)$ is continuous on I. By comparing (56) with Theorem 2.30, it is easy to see that $\Delta(\kappa) > 0$ for all $\kappa > 0$. We can now apply the extreme value theorem, which implies that Δ achieves its minimum on I at some fixed point $\kappa^* \in I$. We therefore have that $\Delta(\kappa) \ge \Delta(\kappa^*) > 0$ for all $\kappa \in I$.

Combining the last two paragraphs, we now know that there is some value

$$\delta := \min\{0.5, \Delta(\kappa^*)\} > 0$$

such that $\Delta(\kappa) > \delta$ for all $\kappa \in (0, \sqrt{C}]$. Crucially, δ is an absolute constant which depends only on the constant C and is independent of n and d. Now, let us apply Theorem 7.2 with the values $f(d) = \sqrt{d}$, C, and $\frac{\delta}{2}$. Then with probability at least $1 - \frac{\delta}{2}$, $d_1(\overline{\boldsymbol{\lambda}}, \Omega_\theta) < \frac{\delta}{2}$. When this occurs,

$$d_{\mathrm{TV}}(\underline{\boldsymbol{\lambda}}, \mathsf{Unif}_d) = \frac{1}{4} d_1(\overline{\boldsymbol{\lambda}}, \overline{\mathsf{unif}_\theta}) \ge \frac{1}{4} \left(d_1(\Omega_\theta, \overline{\mathsf{unif}_\theta}) - d_1(\overline{\boldsymbol{\lambda}}, \Omega_\theta) \right) \ge \frac{\delta}{8},$$

where the second step follows from the triangle inequality, and the third step uses the fact that $d_1(\Omega_\theta, \overline{\mathsf{unif}_\theta}) = \Delta(\theta) \ge \delta$. This proves the theorem with the parameters $1 - \frac{\delta}{2}$ and $\frac{\delta}{8}$. □

It remains to prove Theorem 7.2, and this is done in the next subsection.

7.1 Proof of Theorem 7.2

Our goal is to give a rate of convergence of $\overline{\boldsymbol{\lambda}}$ to Ω_θ which depends only on d and is independent of n. To do this, we will show that standard law of large numbers arguments give convergence rates of this form. Biane's [Bia01] proof of the law of large numbers for the Schur-Weyl distribution does not use Kerov's algebra of observables. Instead, we will follow the proof of the law of large numbers (second form) for the Plancherel distribution in [IO02, Theorem 5.5] and use results from [Mél10a] to extend this proof to the Schur–Weyl distribution. We emphasize that our proof contains no ideas not already found in [IO02, Mél10a], and that our goal is just to show that proper bookkeeping of their arguments yields our Theorem 7.2. (Finally, we note that Meliot [Mél10a] also sketches a proof the law of large numbers for the Schur–Weyl distribution using Kerov's algebra of observables at the beginning of his Section 3.)

Write $\Delta_{\boldsymbol{\lambda}}(x) := \overline{\boldsymbol{\lambda}}(x) - \Omega_\theta(x)$. Because $\overline{\boldsymbol{\lambda}}$ and Ω_θ are both continual diagrams, we know that $\Delta_{\boldsymbol{\lambda}}$ is supported (i.e., nonzero) on a finite interval. We will need a stronger property, which is that the width of this interval does not grow with d (or, equivalently, with n). To show this, note that $\Delta_{\boldsymbol{\lambda}}(x)$ is zero when both $\Omega_\theta(x) = |x|$ and $\overline{\boldsymbol{\lambda}}(x) = |x|$. For the first of these, we can consult Theorem 2.30 and see that $\Omega_\theta(x) = |x|$ outside the interval $[-2, \theta + 2]$. On the other hand, $\overline{\boldsymbol{\lambda}}(x)$ does not equal $|x|$ outside a constant-width interval for all $\boldsymbol{\lambda} \sim \mathrm{SW}_d^n$. (For example, with nonzero probability $\boldsymbol{\lambda} = (n)$, in which case $\overline{\boldsymbol{\lambda}}(x) = |x|$ only outside the interval $(-1/\sqrt{n}, \sqrt{n})$.) However, the next proposition shows that our desired property occurs with high probability.

Proposition 7.3. *With probability* $1 - \frac{\delta}{2}$, $\overline{\boldsymbol{\lambda}}(x) \neq |x|$ *only on an interval of width* $w = O_\delta(1)$.

Proof. We will show that $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}'_1 \leq \beta\sqrt{n}$, each with probability $1 - \delta/4$, for some constant β which depends only on δ (and C). The proposition will then follow from the union bound, as $\overline{\boldsymbol{\lambda}} = |x|$ outside the interval $[-\boldsymbol{\lambda}_1/\sqrt{n}, \boldsymbol{\lambda}_1/\sqrt{n}]$. By Proposition 2.31,

$$\mathbf{Pr}[\boldsymbol{\lambda}_1 \geq \beta\sqrt{n}], \mathbf{Pr}[\boldsymbol{\lambda}'_1 \geq \beta\sqrt{n}] \leq \left(\frac{(1+\beta\theta)e^2}{\beta^2}\right)^{\beta\sqrt{n}} \leq \frac{(1+\beta\theta)e^2}{\beta^2} \leq \frac{(1+\beta\sqrt{C})e^2}{\beta^2}.$$

This can be made less than $\delta/4$ by choosing β to be a sufficiently large function of C and δ. □

Let I' be the constant-width interval guaranteed by Proposition 7.3. Clearly, I' contains the point zero. Thus, if we define

$$I := [-2, \theta + 2] \cup I'$$

then this is a single interval of width $w = O_\delta(1)$. This motivates the following definition:

Definition 7.4. We say that λ is *usual* if Δ_λ is supported on I. By the previous discussion, a random $\boldsymbol{\lambda}$ is usual with probability $1 - \delta/2$.

Let us condition $\boldsymbol{\lambda}$ on it being usual, and let us suppose that $d_1(\overline{\boldsymbol{\lambda}}, \Omega_\theta) \geq \delta$. Then there is some point $x \in I$ such that $|\Delta_{\boldsymbol{\lambda}}(x)| \geq \frac{\delta}{w}$. Now we will use the fact that Ω_θ and $\overline{\boldsymbol{\lambda}}$ are continual diagrams, which implies that they are both 1-Lipschitz, and therefore $\Delta_{\boldsymbol{\lambda}}$ is 2-Lipschitz. Then if we consider the subinterval $I_x \subseteq I$ defined as $I_x := [x - \frac{\delta}{4w}, x + \frac{\delta}{4w}]$, this Lipschitz property implies that $|\Delta_{\boldsymbol{\lambda}}(y)| \geq \frac{\delta}{2w}$ for all $y \in I_x$. (That I_x is contained in I follows from the fact that $\Delta_{\boldsymbol{\lambda}}$ is nonzero on I_x and $\boldsymbol{\lambda}$ is usual.) We note that the width of I_x is $\frac{\delta}{2w}$.

Let $\mathcal{J}$ be a set of $\lceil \frac{4w^2}{\delta} \rceil$ closed intervals of width $\frac{\delta}{4w}$ which cover I. These intervals are chosen to have half the width of I_x, the result being that there is some interval $J^* \in \mathcal{J}$ which is completely

contained in I_x. For each interval $J \in \mathcal{J}$, let $\Psi_J : \mathbb{R} \to \mathbb{R}^{\geq 0}$ be a continuous function supported on J which satisfies $\int \Psi_J(y)dy = 1$ (such functions are known to exist; e.g., bump functions). Then

$$\left|\int_{-\infty}^{\infty} \Delta_{\boldsymbol{\lambda}}(y)\Psi_{J^*}(y)dy\right| \geq \min_{y\in I_x} |\Delta_{\boldsymbol{\lambda}}(y)| \cdot \int_{-\infty}^{\infty} \Psi_{J^*}(y)dy \geq \frac{\delta}{2w}.$$

By the Weierstrass approximation theorem, we can approximate each Ψ_J with a polynomial function $\widetilde{\Psi}_J$ such that for each $x \in I$, $|\Psi_J(x) - \widetilde{\Psi}_J(x)| \leq \frac{\delta}{8w^3}$. (Outside of I, $\widetilde{\Psi}_J$ can—and will—be an arbitrarily bad approximator for Ψ_J.) Because $\Delta_{\boldsymbol{\lambda}}$ is 2-Lipschitz and $\boldsymbol{\lambda}$ is usual, $|\Delta_{\boldsymbol{\lambda}}(x)| \leq 2w$ for all $x \in I$ and is zero everywhere else. As a result, for the interval J^*,

$$\left|\int_{-\infty}^{\infty} \Delta_{\boldsymbol{\lambda}}(y)\widetilde{\Psi}_{J^*}(y)dy\right| \geq \left|\int_{-\infty}^{\infty} \Delta_{\boldsymbol{\lambda}}(y)\Psi_{J^*}(y)dy\right| - \left|\int_{-\infty}^{\infty} \Delta_{\boldsymbol{\lambda}}(y)\left(\Psi_{J^*}(y) - \widetilde{\Psi}_{J^*}(y)\right)dy\right| \geq \frac{\delta}{4w}.$$

The first inequality uses the triangle inequality, and the second inequality uses crucially the fact that $\Delta_{\boldsymbol{\lambda}}$ is zero outside I.

In summary, we have

$$\mathop{\mathbf{Pr}}_{\boldsymbol{\lambda}\sim \mathrm{SW}_d^n}\left[d_1(\overline{\boldsymbol{\lambda}}, \Omega_\theta) \geq \delta\right] \leq \mathop{\mathbf{Pr}}_{\boldsymbol{\lambda}\sim \mathrm{SW}_d^n}\left[\exists J \in \mathcal{J} : \left|\int_{-\infty}^{\infty} \Delta_{\boldsymbol{\lambda}}(y)\widetilde{\Psi}_J(y)dy\right| \geq \frac{\delta}{4w}\right] + \frac{\delta}{2}, \tag{57}$$

where the $\delta/2$ comes from the event that $\boldsymbol{\lambda}$ is not usual. We will therefore show that $\left|\int \Delta_{\boldsymbol{\lambda}}(y)\widetilde{\Psi}_J(y)dy\right|$ is at most $\frac{\delta}{4w}$ for all $J \in \mathcal{J}$ with probability at least $1 - \frac{\delta}{2}$. By the union bound, it suffices to show that for each $J \in \mathcal{J}$, $\left|\int \Delta_{\boldsymbol{\lambda}}(y)\widetilde{\Psi}_J(y)dy\right| \leq \frac{\delta}{4w}$ with probability at least $1 - \frac{\delta}{2\cdot|\mathcal{J}|}$.

Let m be the maximum degree of the $\widetilde{\Psi}_J$ functions, for all $J \in \mathcal{J}$. Fix an interval $J \in \mathcal{J}$. Then we can write

$$\widetilde{\Psi}_J(x) = \sum_{k=0}^{m} a_J^{(k)} x^k \quad \text{and} \quad \int_{-\infty}^{\infty} \Delta_\lambda(y)\widetilde{\Psi}_J(y)dy = \sum_{k=0}^{m} a_J^{(k)} \int_{-\infty}^{\infty} x^k \Delta_\lambda(x)dx, \tag{58}$$

where the $a_J^{(k)}$'s are constants. The following proposition, found in [Mél10a, Lemma 7], gives a nice expression for the integrals on the right-hand side.

Proposition 7.5. *Let $k \geq 1$. Then*

$$\int_{-\infty}^{\infty} x^k \Delta_\lambda(x)dx = \frac{2 \cdot \widetilde{q}_{k+1}(\lambda)}{(k+1)\sqrt{n}},$$

where $\widetilde{q}_k(\lambda)$ is the quantity defined as

$$\widetilde{q}_k(\lambda) \coloneqq \frac{\widetilde{p}_{k+1}(\lambda)}{(k+1)n^{k/2}} - \sum_{\ell=1}^{\lfloor \frac{k+1}{2} \rfloor} \frac{k^{\downarrow 2\ell-1}}{(k+1-\ell)\ell!(\ell-1)!} \cdot \frac{n^{k/2+1-\ell}}{d^{k+1-2\ell}}.$$

The key fact we will use is that we can upper bound the right-hand side of Equation (58) by a quantity which decays with d, independent of the value of n. This is the subject of the following lemma.

Lemma 7.6. *The random variable $\left|\frac{\widetilde{q}_k(\boldsymbol{\lambda})}{\sqrt{n}}\right|$, for $\boldsymbol{\lambda} \sim \mathrm{SW}_d^n$, has mean $o_d(1)$, for all $f(d) \leq n \leq Cd^2$.*

Applying Proposition 7.5 and Lemma 7.6 to Equation (58), we see that $\mathbf{E}_{\boldsymbol{\lambda}\sim \mathrm{SW}_d^n}\left|\int \Delta_{\boldsymbol{\lambda}}(x)\widetilde{\Psi}_J(x)dx\right|$ is $o_d(1)$. We may take d large enough to make this quantity arbitrarily small. Thus, select d_J so that for all $d \geq d_J$, this expectation is at most $\frac{\delta^2}{8w\cdot|\mathcal{J}|}$. Then by Markov's inequality, $\left|\int \Delta_{\boldsymbol{\lambda}}(x)\widetilde{\Psi}_J(x)dx\right| \leq \frac{\delta}{4w}$ with probability at least $1-\frac{\delta}{2\cdot|\mathcal{J}|}$. If we set d_0 to be the max of d_J over all $J\in\mathcal{J}$, then by Equation (57), $\mathbf{Pr}_{\boldsymbol{\lambda}\sim \mathrm{SW}_d^n}\left[d_1(\overline{\boldsymbol{\lambda}},\Omega_\theta)\geq\delta\right]\leq\delta$ so long as $d\geq d_0$, and we are done.

Now we turn to the proof of Lemma 7.6.

Proof of Lemma 7.6. Define

$$X_k(\lambda) \coloneqq \sum_{\mu:\mathrm{wt}(\mu)=k} \frac{k^{\downarrow\ell(\mu)}}{m(\mu)}\cdot p_\mu^\sharp(\lambda)$$

and

$$q_k^\sharp(\lambda) \coloneqq \frac{X_{k+1}(\lambda)}{(k+1)n^{k/2}} - \sum_{\ell=1}^{\lfloor\frac{k+1}{2}\rfloor} \frac{k^{\downarrow 2\ell-1}}{(k+1-\ell)\ell!(\ell-1)!}\cdot\frac{n^{k/2+1-\ell}}{d^{k+1-2\ell}}. \tag{59}$$

Then by Proposition 5.5, $\widetilde{q}_k(\lambda)$ and $q_k^\sharp(\lambda)$ differ from each other by $n^{-k/2}$ times an observable $\mathcal{O}(\lambda)$ of weight k. Thus,

$$\mathop{\mathbf{E}}_{\boldsymbol{\lambda}\sim \mathrm{SW}_d^n}\left|\frac{\widetilde{q}_k(\boldsymbol{\lambda})}{\sqrt{n}}\right| \leq \mathop{\mathbf{E}}_{\boldsymbol{\lambda}\sim \mathrm{SW}_d^n}\left|\frac{q_k^\sharp(\boldsymbol{\lambda})}{\sqrt{n}}\right| + \mathop{\mathbf{E}}_{\boldsymbol{\lambda}\sim \mathrm{SW}_d^n}\left|\frac{\mathcal{O}(\boldsymbol{\lambda})}{n^{(k+1)/2}}\right|.$$

By Cauchy–Schwarz, $\mathbf{E}\,|\mathcal{O}(\boldsymbol{\lambda})/n^{(k+1)/2}| \leq \sqrt{\mathbf{E}\,\mathcal{O}(\boldsymbol{\lambda})^2/n^{k+1}}$. Because $\mathcal{O}$ has weight k, $\mathcal{O}^2$ has weight $2k$. As a result, we can use the next proposition to bound the contribution from this term by $o_d(1)$.

Proposition 7.7. *Let $\mathcal{O}(\lambda)$ be an observable of weight at most $2k$. Then*

$$\mathop{\mathbf{E}}_{\boldsymbol{\lambda}\sim \mathrm{SW}_d^n}\left[\frac{\mathcal{O}(\boldsymbol{\lambda})}{n^{k+1}}\right] = o_d(1).$$

Proof. As in the proof of Lemma 5.7, this reduces to showing that $\mathbf{E}_{\boldsymbol{\lambda}\sim \mathrm{SW}_d^n}\left[p_\mu^\sharp(\boldsymbol{\lambda})/n^{k+1}\right] = o_d(1)$, where μ is a partition of weight $2k$, i.e. $|\mu|+\ell(\mu)\leq 2k$. By Corollary 2.35,

$$\mathop{\mathbf{E}}_{\boldsymbol{\lambda}\sim \mathrm{SW}_d^n}\left[\frac{p_\mu^\sharp(\boldsymbol{\lambda})}{n^{k+1}}\right] = \frac{n^{\downarrow|\mu|}}{n^{k+1}}\cdot\frac{d^{\ell(\mu)}}{d^{|\mu|}} \leq \frac{n^{|\mu|}}{n^{k+1}}\cdot\frac{d^{\ell(\mu)}}{d^{|\mu|}} = \frac{n^{|\mu|}}{n^{k+1}}\cdot\frac{d^{\mathrm{wt}(\mu)}}{d^{2|\mu|}}.$$

If $|\mu|<k+1$, then this expression is at most $1/n$, which is $o_d(1)$ because $n\geq f(d)=\omega_d(1)$. On the other hand, if $|\mu|\geq k+1$, then for all $n\leq Cd^2$ this expression is at most

$$\frac{(Cd^2)^{|\mu|}}{(Cd^2)^{k+1}}\cdot\frac{d^{\mathrm{wt}(\mu)}}{d^{2|\mu|}} \leq C^{|\mu|-(k+1)}\cdot\frac{d^{\mathrm{wt}(\mu)}}{d^{2(k+1)}},$$

which is $o_d(1)$ as $\mathrm{wt}(\mu)\leq 2k$. □

It remains to bound $\mathbf{E}\,|q_k^\sharp(\boldsymbol{\lambda})/\sqrt{n}|$ by $o_d(1)$. First, we will show that $q_k^\sharp(\boldsymbol{\lambda})$ can be viewed as (approximately) computing the deviation of a certain random variable from its mean. To do this, let us compute the mean of the first term on the right-hand side of Equation (59).

$$\begin{aligned}
\mathop{\mathbf{E}}_{\boldsymbol{\lambda}\sim \mathrm{SW}_d^n} \frac{X_{k+1}(\boldsymbol{\lambda})}{(k+1)n^{k/2}} &= \frac{1}{(k+1)n^{k/2}} \cdot \sum_{\mu:\mathrm{wt}(\mu)=k+1} \frac{(k+1)^{\downarrow \ell(\mu)}}{m(\mu)} \cdot \frac{n^{\downarrow|\mu|}}{d^{|\mu|-\ell(\mu)}} \\
&= \frac{1}{(k+1)n^{k/2}} \cdot \sum_{\ell=1}^{\lfloor \frac{k+1}{2} \rfloor} \frac{(k+1)^{\downarrow \ell} n^{\downarrow k+1-\ell}}{d^{k+1-2\ell}} \sum_{\mu:\mathrm{wt}(\mu)=k+1} \frac{1}{m(\mu)} \\
&= \frac{1}{(k+1)n^{k/2}} \cdot \sum_{\ell=1}^{\lfloor \frac{k+1}{2} \rfloor} \frac{(k+1)^{\downarrow \ell} n^{\downarrow k+1-\ell}}{d^{k+1-2\ell}} \cdot \frac{1}{\ell!}\binom{k-\ell}{\ell-1} \\
&= \sum_{\ell=1}^{\lfloor \frac{k+1}{2} \rfloor} \frac{k^{\downarrow 2\ell-1}}{(k+1-\ell)\ell!(\ell-1)!} \cdot \frac{n^{\downarrow k+1-\ell}}{n^{k/2} \cdot d^{k+1-2\ell}},
\end{aligned}$$

where the third equality follows from [Mél10a, Lemma 11]. As a result, the difference

$$\mathop{\mathbf{E}}_{\boldsymbol{\lambda}\sim \mathrm{SW}_d^n} \frac{X_{k+1}(\boldsymbol{\lambda})}{(k+1)n^{k/2}} - \sum_{\ell=1}^{\lfloor \frac{k+1}{2} \rfloor} \frac{k^{\downarrow 2\ell-1}}{(k+1-\ell)\ell!(\ell-1)!} \cdot \frac{n^{k/2+1-\ell}}{d^{k+1-2\ell}}$$

can be written as a sum over terms of the form $a \cdot n^b/d^{k+1-2\ell}$, where a is a constant coefficient, $1 \le b \le k/2-\ell$, and $1 \le \ell \le \lfloor \frac{k+1}{2} \rfloor$. Given that $n \le Cd^2$, each of these terms if $\pm o_d(1)$. Thus, if we set

$$q_k(\lambda) \coloneqq \frac{X_{k+1}(\lambda)}{(k+1)n^{k/2}} - \mathop{\mathbf{E}}_{\boldsymbol{\lambda}\sim \mathrm{SW}_d^n} \frac{X_{k+1}(\boldsymbol{\lambda})}{(k+1)n^{k/2}},$$

then

$$\mathop{\mathbf{E}}_{\boldsymbol{\lambda}\sim \mathrm{SW}_d^n} \left| \frac{q_k^\sharp(\boldsymbol{\lambda})}{\sqrt{n}} \right| \le \mathop{\mathbf{E}}_{\boldsymbol{\lambda}\sim \mathrm{SW}_d^n} \left| \frac{q_k(\boldsymbol{\lambda})}{\sqrt{n}} \right| + o_1(d).$$

Finally, we show that $\mathbf{E}\,|q_k(\boldsymbol{\lambda})/\sqrt{n}| = o_d(1)$. By Cauchy–Schwarz,

$$\mathbf{E} \left| \frac{q_k(\boldsymbol{\lambda})}{\sqrt{n}} \right| \le \sqrt{\mathbf{E} \left(\frac{q_k(\boldsymbol{\lambda})}{\sqrt{n}} \right)^2},$$

so it suffices to show that $\mathbf{E}\,(q_k(\boldsymbol{\lambda})/\sqrt{n})^2 = o_d(1)$. This expectation is simply the variance of the random variable $X_{k+1}(\boldsymbol{\lambda})/(k+1)n^{(k+1)/2}$, which itself is a weighted sum of a constant number of random variables of the form $p_\mu^\sharp(\boldsymbol{\lambda})/n^{(k+1)/2}$, where $\mathrm{wt}(\mu) = k+1$. An easy application of Cauchy–Schwarz shows that the variance of a weighted sum of a constant number of random variables is $o_d(1)$ if the variance of each random variables is $o_d(1)$. Thus, we will show that $\mathbf{Var}[p_\mu^\sharp(\boldsymbol{\lambda})/n^{(k+1)/2}] = o_d(1)$ for all $\mathrm{wt}(\mu) = k+1$.

Fix a partition μ of weight $k+1$. Then

$$\mathbf{Var}\left[\frac{p_\mu^\sharp(\boldsymbol{\lambda})}{n^{(k+1)/2}} \right] = \mathop{\mathbf{E}}_{\boldsymbol{\lambda}\sim \mathrm{SW}_d^n} \left[\frac{1}{n^{(k+1)/2}} \left(p_\mu^\sharp(\boldsymbol{\lambda}) p_\mu^\sharp(\boldsymbol{\lambda}) - \mathbf{E}[p_\mu^\sharp]^2 \right) \right]$$

By Proposition 2.39, $p^\sharp_\mu(\lambda) \cdot p^\sharp_\mu(\lambda) = p^\sharp_{\mu\cup\mu}(\lambda) + \mathcal{O}(\lambda)$, where $\mathcal{O}(\lambda)$ is an observable of weight at most $2 \cdot \mathrm{wt}(p^\sharp_\mu) - 2 = 2k$. Then

$$\mathbf{Var}\left[\frac{p^\sharp_\mu(\boldsymbol{\lambda})}{n^{(k+1)/2}}\right] = \mathop{\mathbf{E}}_{\boldsymbol{\lambda}\sim \mathrm{SW}^n_d}\left[\frac{1}{n^{k+1}} \cdot \left(p^\sharp_{\mu\cup\mu}(\boldsymbol{\lambda}) - \mathbf{E}[p^\sharp_\mu]^2\right)\right] + \mathop{\mathbf{E}}_{\boldsymbol{\lambda}\sim \mathrm{SW}^n_d}\left[\frac{1}{n^{k+1}} \cdot \mathcal{O}(\boldsymbol{\lambda})\right].$$

The second term is $\pm o_d(1)$ by Proposition 7.7. As for the first term, Corollary 2.35, shows that it equals

$$\frac{1}{n^{k+1}} \cdot \left(n^{\downarrow 2|\mu|} d^{2\ell(\mu)-2|\mu|} - n^{\downarrow|\mu|} n^{\downarrow|\mu|} d^{2\ell(\mu)-2|\mu|}\right) = \frac{1}{d^{4|\mu|-2(k+1)}} \cdot \left(\frac{n^{\downarrow 2|\mu|} - (n^{\downarrow|\mu|})^2}{n^{k+1}}\right), \tag{60}$$

where we used the fact that $\ell(\mu) = \mathrm{wt}(\mu) - |\mu| = k + 1 - |\mu|$. The highest-degree term of both $n^{\downarrow 2|\mu|}$ and $(n^{\downarrow|\mu|})^2$ is $n^{2|\mu|}$, so we can write

$$(60) = \frac{1}{d^{4|\mu|-2(k+1)}} \cdot \sum_{b=-(k+1)}^{2|\mu|-(k+2)} \alpha_b \cdot n^b$$

for some constants α_b. When $b < 0$, $n^b/d^{4|\mu|-2k-2} \le 1/n$, which is $o_d(1)$ because $n \ge f(d) = \omega_d(1)$. On the other hand, when $b \ge 0$, then this term is $o_d(1)$ because $n \le Cd^2$. □

References

[ARS88] Robert Alicki, Sławomir Rudnicki, and Sławomir Sadowski. Symmetry properties of product states for the system of N n-level atoms. *Journal of mathematical physics*, 29(5):1158–1162, 1988. (document), 1.2, 1.4, 2.6, 3

[Aud06] Koenraad Audenaert. A digest on representation theory of the symmetric group. Found at http://personal.rhul.ac.uk/usah/080/qitnotes_files/irreps_v06.pdf, 2006. 2.6

[Bat01] Tuğkan Batu. *Testing properties of distributions*. PhD thesis, Cornell University, 2001. 6

[Bay02] Richard Bayley. Young tableaux and the Robinson–Schensted–Knuth correspondence. Master's thesis, University of Leicester, 2002. 2.3.2

[BCH05] Dave Bacon, Isaac Chuang, and Aram Harrow. The quantum Schur transform: I. efficient qudit circuits. In *Proceedings of the 18th Annual ACM-SIAM Symposium on Discrete Algorithms*, 2005. 2.5

[BDKR05] Tuğkan Batu, Sanjoy Dasgupta, Ravi Kumar, and Ronitt Rubinfeld. The complexity of approximating the entropy. *SIAM Journal on Computing*, 35(1):132–150, 2005. 1.1

[BFF+01] Tuğkan Batu, Eldar Fischer, Lance Fortnow, Ravi Kumar, Ronitt Rubinfeld, and Patrick White. Testing random variables for independence and identity. In *Proceedings of the 42nd Annual IEEE Symposium on Foundations of Computer Science*, pages 442–451, 2001. 1.1

[BFR+00] Tuğkan Batu, Lance Fortnow, Ronitt Rubinfeld, Warren Smith, and Patrick White. Testing that distributions are close. In *Proceedings of the 41st Annual IEEE Symposium on Foundations of Computer Science*, pages 259–269, 2000. 1.1, 1.1

[BFR+13] Tuğkan Batu, Lance Fortnow, Ronitt Rubinfeld, Warren Smith, and Patrick White. Testing closeness of discrete distributions. *Journal of the ACM*, 60(1):4, 2013. 1.1, 1.1

[Bia01] Philippe Biane. Approximate factorization and concentration for characters of symmetric groups. *International Mathematics Research Notices*, 2001(4):179–192, 2001. 1.4, 2.7, 2.30, 7.1

[Bro] Daniel Brown. How I wasted too long finding a concentration inequality for sums of geometric variables. Found at https://cs.uwaterloo.ca/~browndg/negbin.pdf. 6.3

[CGS04] Sylvie Corteel, Alain Goupil, and Gilles Schaeffer. Content evaluation and class symmetric functions. *Advances in Mathematics*, 188(2):315–336, 2004. 5.2.1

[Chr06] Matthias Christandl. *The Structure of Bipartite Quantum States*. PhD thesis, University of Cambridge, 2006. 3.1

[CHW07] Andrew Childs, Aram Harrow, and Paweł Wocjan. Weak Fourier-Schur sampling, the hidden subgroup problem, and the quantum collision problem. In *24th Annual Symposium on Theoretical Aspects of Computer Science*, pages 598–609, 2007. 1, 1.2, 1.4, 1.4, 2.5, 2.5, 2.6, 2.6, 2.7, 4.1, 4.2

[CM06] Matthias Christandl and Graeme Mitchison. The spectra of quantum states and the Kronecker coefficients of the symmetric group. *Communications in mathematical physics*, 261(3):789–797, 2006. (document), 1.2, 1.4, 3.1, 3.1

[CSST10] Tullio Ceccherini-Silberstein, Fabio Scarabotti, and Filippo Tolli. *Representation theory of the symmetric groups: the Okounkov-Vershik approach, character formulas, and partition algebras*. Cambridge University Press, 2010. 5.2.1

[Dia14] Ilias Diakonikolas. Beyond histograms: structure and distribution estimation. Found at http://www.iliasdiakonikolas.org/stoc14-workshop/diakonikolas.pdf, 2014. 1.1

[DL01] Luc Devroye and Gábor Lugosi. *Combinatorial methods in density estimation*. Springer, 2001. 1.1

[Dor05] Heather Dornom. Robinson–Schensted–Knuth correspondence. Master's thesis, University of Melbourne, 2005. 2.3.2

[Fér10] Valentin Féray. Stanley's formula for characters of the symmetric group. *Annals of Combinatorics*, 13(4):453–461, 2010. 2.32

[FGLE12] Steven Flammia, David Gross, Yi-Kai Liu, and Jens Eisert. Quantum tomography via compressed sensing: error bounds, sample complexity and efficient estimators. *New Journal of Physics*, 14(9):095022, 2012. 1, 1.3

[FRT54] James Frame, Gilbert Robinson, and Robert Thrall. The hook graphs of the symmetric group. *Canadian Journal of Mathematics*, 6:316–324, 1954. 2.9

[GKP94] Ronald Graham, Donald Knuth, and Oren Patashnik. *Concrete mathematics: a foundation for computer science*. Addison–Wesley, second edition, 1994. 5.2.4

[GR11] Oded Goldreich and Dana Ron. On testing expansion in bounded-degree graphs. In *Studies in Complexity and Cryptography. Miscellanea on the Interplay between Randomness and Computation*, pages 68–75. Springer, 2011. 1.1

[Gre74] Curtis Greene. An extension of Schensted's theorem. *Advances in Mathematics*, 14:254–265, 1974. 2.3.2

[Ham72] John Hammersley. A few seedlings of research. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, pages 345–394, 1972. 2.7

[Har05] Aram Harrow. *Applications of coherent classical communication and the Schur transform to quantum information theory*. PhD thesis, Massachusetts Institute of Technology, 2005. 1.4, 2.5, 2.5, 2.5

[HJ13] Roger Horn and Charles Johnson. *Matrix analysis*. Cambridge University Press, 2nd edition, 2013. 2.2

[HM02] Masahito Hayashi and Keiji Matsumoto. Quantum universal variable-length source coding. *Physical Review A*, 66(2):022311, 2002. (document), 1.2, 3.1, 3.1

[HR18] G. H. Hardy and Srinivasa Ramanujan. Asymptotic formulae in combinatory analysis. *Proceddings of the London Mathematical Society*, 2(17):75–115, 1918. 2.3

[HX13] Christian Houdré and Hua Xu. On the limiting shape of Young diagrams associated with inhomogeneous random words. In *High Dimensional Probability VI*, volume 66 of *Progress in Probability*, pages 277–302. Springer Basel, 2013. 3

[IK01] Vladimir Ivanov and Sergei Kerov. The algebra of conjugacy classes in symmetric groups and partial permutations. *Journal of Mathematical Sciences*, 107(5):4212–4230, 2001. 2.32, 2.8.1, 2.8.1

[IO02] Vladimir Ivanov and Grigori Olshanski. Kerov's central limit theorem for the Plancherel measure on Young diagrams. In *Symmetric functions 2001: surveys of developments and perspectives*, pages 93–151. Springer, 2002. 1.4, 2.7, 2.32, 2.8.1, 2.8.1, 2.8.1, 5.2.1, 5.2.2, 7.1

[ITW01] Alexander Its, Craig Tracy, and Harold Widom. Random words, Toeplitz determinants and integrable systems I. In *Random Matrices and their Applications*, pages 245–258. Cambridge University Press, 2001. 2.3.2, 3, 3.2

[JK81] Gordon James and Adalbert Kerber. *The representation theory of the symmetric group*. Addison–Wesley, 1981. 4.13

[Joh01] Kurt Johansson. Discrete orthogonal polynomial ensembles and the Plancherel measure. *Annals of Mathematics*, 153(1):259–296, 2001. 1.4

[Ker93a] Sergei Kerov. The asymptotics of root separation for orthogonal polynomials. *Algebra i Analiz*, 5(5):68–86, 1993. 2.7

[Ker93b] Sergei Kerov. Gaussian limit for the Plancherel measure of the symmetric group. *Comptes Rendus de l'Académie des Sciences, Série 1*, 316:303–308, 1993. 2.7

[Knu70] Donald Knuth. Permutations, matrices, and generalized Young tableaux. *Pacific Journal of Mathematics*, 34(3):709–727, 1970. 2.3.2, 2.3.2

[KO94] Sergei Kerov and Grigori Olshanski. Polynomial functions on the set of Young diagrams. *Comptes Rendus de l'Académie des Sciences, Série 1*, 319(2):121–126, 1994. 2.32

[Kup02] Greg Kuperberg. Random words, quantum statistics, central limits, random matrices. *Methods and Applications of Analysis*, 9(1):99–118, 2002. 1.4

[KW01] Michael Keyl and Reinhard Werner. Estimating the spectrum of a density operator. *Physical Review A*, 64(5):052311, 2001. (document), 1.2, 3, 3.1

[Las78] Alain Lascoux. Classes de chern d'un produit tensoriel. *Comptes Rendus de l'Académie des Sciences, Série 1*, 286:385–387, 1978. 4.3

[Las08] Michel Lassalle. An explicit formula for the characters of the symmetric group. *Mathematische Annalen*, 340(2):383–405, 2008. 2.32, 5.2.1

[Lit08] Trevis Litherland. *On the limiting shape of random young tableaux for Markovian words*. PhD thesis, Georgia Institute of Technology, 2008. 3

[LS77] Benjamin Logan and Larry Shepp. A variational problem for random Young tableaux. *Advances in Mathematics*, 26(2):206–222, 1977. 2.7, 2.7

[Mac95] Ian Macdonald. *Symmetric functions and Hall polynomials*. Oxford University Press, 1995. 8, 2.32, 4.3, 5.2.1

[MdW13] Ashley Montanaro and Ronald de Wolf. A survey of quantum property testing. Technical report, arXiv:1310.2035, 2013. (document), 1, 1.1, 1.4, 2.5

[Mél10a] Pierre-Loïc Méliot. Kerov's central limit theorem for Schur-Weyl measures of parameter 1/2. Technical report, arXiv:1009.4034, 2010. 1.4, 2.7, 5.2.4, 7.1, 7.1, 7.1

[Mél10b] Pierre-Loïc Méliot. *Partitions aléatoires et théorie asymptotique des groupes symétriques, des algèbres d'Hecke et des groupes de Chevalley finis*. PhD thesis, University Paris-Est Marne-la-Vallée, 2010. 2.7, 2.32, 5.2.2

[Mol09] Alexander Molev. Littlewood-Richardson polynomials. *Journal of Algebra*, 321(11):3450–3468, 2009. 4.3

[Mon14] Ashley Montanaro. Personal communication, 2014. 2.2

[MS99] Alexander Molev and Bruce Sagan. A Littlewood-Richardson rule for factorial Schur functions. *Transactions of the American Mathematical Society*, 351(11):4429–4443, 1999. 4.3

[NC10] Michael Nielsen and Isaac Chuang. *Quantum computation and quantum information*. Cambridge university press, 2010. 2.1

[OO98a] Andrei Okounkov and Grigori Olshanski. Asymptotics of Jack polynomials as the number of variables goes to infinity. *International Mathematics Research Notices*, 13:641–682, 1998. 4.3

[OO98b] Andrei Okounkov and Grigori Olshanski. Shifted Schur functions. *St. Petersburg Mathematical Journal*, 9(2):239–300, 1998. 2.3, 2.8, 2.32, 4.3

[Pan04] Liam Paninski. Estimating entropy on m bins given fewer than m samples. *IEEE Transactions on Information Theory*, 50(9):2200–2203, 2004. 1.1

[Pan08] Liam Paninski. A coincidence-based test for uniformity given very sparsely sampled discrete data. *IEEE Transactions on Information Theory*, 54(10):4750–4755, 2008. (document), 1.1, 1.4, 4.2

[PT27] T. E. Phipps and J. B. Taylor. The magnetic moment of the hydrogen atom. *Physical Review*, 29(2):309, 1927. 1

[Rom14] Dan Romik. *The surprising mathematics of longest increasing subsequences*. Cambridge University Press, 2014. 2.3, 2.3.2, 2.7

[RS92] Ronitt Rubinfeld and Madhu Sudan. Self-testing polynomial functions efficiently and over rational domains. In *Proceedings of the 3rd Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 23–32, 1992. 1.1

[RS96] Ronitt Rubinfeld and Madhu Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM Journal on Computing*, 25(2):252–271, 1996. 1.1

[RSW04] Victor Reiner, Dennis Stanton, and Dennis White. The cyclic sieving phenomenon. *Journal of Combinatorial Theory. Series A*, 108(1):17–50, 2004. 4.4

[RZ12] Rodolfo Ríos-Zertuche. *Near-involutions, the pillowcase distribution, and quadratic differentials*. PhD thesis, Princeton University, 2012. 4.15, 4.4, 4.4

[Sag01] Bruce E Sagan. *The symmetric group: representations, combinatorial algorithms, and symmetric functions*. Springer, 2001. 2.19

[Śni06] Piotr Śniady. Asymptotics of characters of symmetric groups, genus expansion and free probability. *Discrete mathematics*, 306(7):624–665, 2006. 2.8.1

[Sta99] Richard P Stanley. *Enumerative combinatorics Volume 2*. Cambridge University Press, Cambridge, 1999. 2.9, 2.10, 2.3.1, 2.4

[Sta11] Richard P Stanley. *Enumerative combinatorics Volume 1*. Cambridge University Press, Cambridge, 2011. 2.7

[TW01] Craig Tracy and Harold Widom. On the distributions of the lengths of the longest monotone subsequences in random words. *Probability Theory and Related Fields*, 119(3):350–380, 2001. 1.4

[Val08] Paul Valiant. *Testing symmetric properties of distributions*. PhD thesis, Massachusetts Institute of Technology, 2008. 1.1, 6

[VK77] Anatoly Vershik and Sergei Kerov. Asymptotic behavior of the Plancherel measure of the symmetric group and the limit form of Young tableaux. *Soviet Mathematics Doklady*, 18:118–121, 1977. 2.7, 2.7

[VK81] Anatoly Vershik and Sergei Kerov. Asymptotic theory of characters of the symmetric group. *Functional analysis and its applications*, 15(4):246–255, 1981. 2.32

[VV11a] Gregory Valiant and Paul Valiant. Estimating the unseen: an $n/\log(n)$-sample estimator for entropy and support size, shown optimal via new CLTs. In *Proceedings of the 43rd Annual ACM Symposium on Theory of Computing*, pages 685–694, 2011. 1.1, 1.1

[VV11b] Gregory Valiant and Paul Valiant. The power of linear estimators. In *Proceedings of the 52nd Annual IEEE Symposium on Foundations of Computer Science*, pages 403–412, 2011. 1.1

[VV14] Gregory Valiant and Paul Valiant. An automatic inequality prover and instance optimal identity testing. In *Proceedings of the 55th Annual IEEE Symposium on Foundations of Computer Science*, 2014. 1.1

[Was81] Antony John Wassermann. *Automorphic actions of compact groups on operator algebras*. PhD thesis, University of Pennsylvania, 1981. 2.32, 2.8.1